

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



21 APR 2005



(43) International Publication Date
6 May 2004 (06.05.2004)

PCT

(10) International Publication Number
WO 2004/037982 A2

- (51) International Patent Classification⁷: C12N
- (21) International Application Number:
PCT/US2003/033433
- (22) International Filing Date: 22 October 2003 (22.10.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/421,252 24 October 2002 (24.10.2002) US
- (71) Applicant (for all designated States except US): BIO-
GEN, INC. [US/US]; 14 Cambridge Center, Cambridge,
MA 02142 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): PRENTICE, Holly
[US/US]; 35 Nathan Lane, Carlisle, MA 01741 (US).
- (74) Agent: FRASER, Janis, K., J., D.; Fish & Richardson
P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

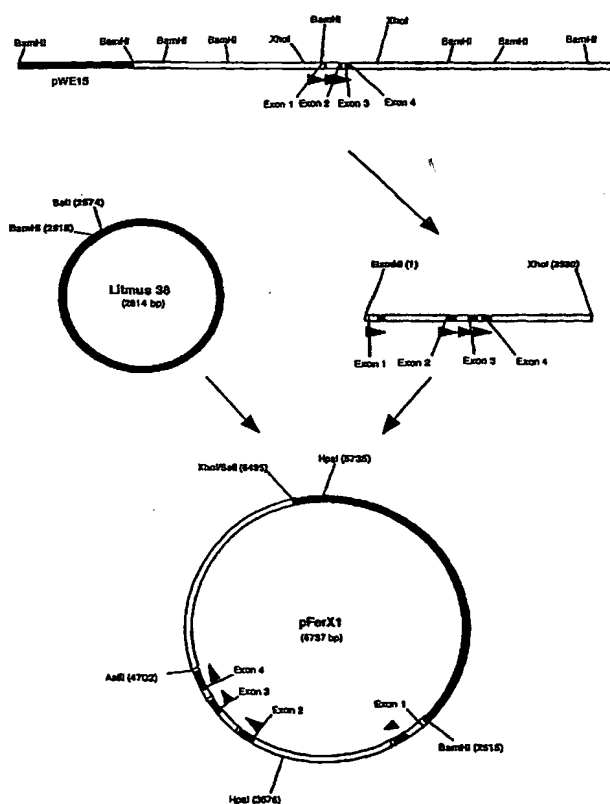
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE,
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT,
RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,
SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted
a patent (Rule 4.17(ii)) for the following designations AE,
AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH,

[Continued on next page]

(54) Title: HIGH EXPRESSION LOCUS VECTOR BASED ON FERRITIN HEAVY CHAIN GENE LOCUS



(57) Abstract: High expression locus vectors based, in part, on the ferritin heavy chain locus are disclosed. The vectors include distal 5' flanking sequences and/or proximal 5' regulatory sequences derived from ferritin heavy chain locus. The vectors include a site for insertion of heterologous sequences and proximal 3' regulatory and distal 3' flanking sequences. The proximal 3' regulatory and distal 3' flanking sequences are optionally derived from the ferritin heavy chain locus. Cells transformed with the vectors, and methods of producing heterologous proteins encoded by the vectors, are also disclosed.



CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU,

LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

HIGH EXPRESSION LOCUS VECTOR BASED ON FERRITIN HEAVY CHAIN GENE LOCUS

5 BACKGROUND OF THE INVENTION

Field of the Invention.

[0001] This invention relates to the field of molecular biology, and in particular to the development and use of vectors for the expression of heterologous genetic sequences in transformed cells.

10 Description of the Related Art.

[0002] Typical expression vectors contain promoters to drive the gene of interest as well as polyadenylation signals to generate a mature transcript. Promoter sequences tend to be only a few hundred base pairs in length and contain most, if not all, of the regulatory regions for optimal expression as determined by transient
15 transfection. However, expression constructs containing these sequences, although highly functional in transient transfections, are not always able to confer a similar level of expression when integrated into the chromatin as a stable transfectant. This is due to position-dependent expression, a phenomenon in which the site of integration has a dominant effect, usually negative, on the level of expression (Wilson (1990), *Ann. Rev.*
20 *Cell Biol.* 6:679-714). The result of position-dependent expression is evident in the results of a transfection screening, in which most of the cell lines produce little or no product. Therefore, it is usually necessary to screen a large number of transfectants in order to identify a single high-expressing clone. Even after extensive screening, transfectants obtained using standard expression vectors typically have expression
25 levels that would not be sufficient to meet commercial titer goals.

[0003] The time consuming and labor intensive process of DHFR amplification is frequently employed to increase expression levels in stable transfectants. For example, integrated copies of standard expression constructs typically require amplification to greater than 100 copies in order to approach the level
30 of expression of endogenous genes with promoters of similar strength (from only two alleles). The differences between standard expression vectors and endogenous genes are most likely due to the presence of sequences 5' to the promoter and/or 3' to the

are most likely due to the presence of sequences 5' to the promoter and/or 3' to the polyadenylation signal of the endogenous genes that are able to confer a chromatin configuration more favourable for expression. An expression construct containing sequences that can confer favourable position-independent chromatin configurations, regardless of the integration site(s) would be advantageous for generating cell lines highly expressing heterologous genes.

SUMMARY OF THE INVENTION

[0004] The present invention depends, in part, upon the development of high expression "locus vectors" derived from the ferritin heavy chain gene. The concept of a "locus vector" is based on the observation that the regions found 5' and 3' to highly expressed genes in their natural chromatin contexts can confer higher levels of expression to a heterologous gene. Therefore, the present invention provides ferritin heavy chain gene locus vectors which include 5' and 3' sequences which can convey high levels of expression to heterologous genes in stable transfectants. Thus, the invention provides genetic vectors for the stable transfection and expression at high levels of a desired protein within eukaryotic cells.

[0005] In one aspect, the invention provides genetic vectors for stable transfection and expression of a desired protein within eukaryotic cells including:

(a) distal 5' flanking sequences of a eukaryotic locus; (b) proximal 5' regulatory sequences of a eukaryotic locus; (c) at least a first insertion site for a heterologous sequence; and (d) proximal 3' regulatory sequences effective for transcription termination of a eukaryotic locus; in which these sequences are operably joined in the order (a)-(d) in a 5' to 3' orientation, with optional linker sequences between adjacent sequences; and in which (1) the distal 5' flanking sequences comprise a sequence of at least 100 bases having at least 70% identity to a nucleotide sequence found between 20 bp and 100,000 bp 5' of a transcriptional initiation site of a ferritin heavy chain locus; and/or (2) the proximal 5' regulatory sequences comprise a sequence of at least 20 bases having at least 70% identity to a nucleotide sequence found between 1 bp and 10,000 bp 5' of a translational initiation codon of a ferritin heavy chain locus.

[0006] In another aspect, the vector includes at least a first heterologous coding sequence encoding a desired protein. Thus, the invention provides genetic

vectors for stable transfection and expression of a desired protein within eukaryotic cells including: (a) distal 5' flanking sequences of a eukaryotic locus; (b) proximal 5' regulatory sequences of a eukaryotic locus; (c) at least a first heterologous coding sequence encoding said desired protein; and (d) proximal 3' regulatory sequences effective for transcription termination of a eukaryotic locus; in which these sequences are operably joined in the order (a)-(d) in a 5' to 3' orientation, with optional linker sequences between adjacent sequences; and in which (1) the distal 5' flanking sequences comprise a sequence of at least 100 bases having at least 70% identity to a nucleotide sequence found between 20 bp and 100,000 bp 5' of a transcriptional initiation site of a ferritin heavy chain locus; and/or (2) the proximal 5' regulatory sequences comprise a sequence of at least 20 bases having at least 70% identity to a nucleotide sequence found between 1 bp and 10,000 bp 5' of a translational initiation codon of a ferritin heavy chain locus.

[0007] In some embodiments, the distal 5' flanking sequences are derived from a ferritin heavy chain locus. In other embodiments, the proximal 5' regulatory sequences are derived from a ferritin heavy chain locus. In yet other embodiments, both the proximal 5' regulatory sequences and the distal 5' flanking sequences are derived from a ferritin heavy chain locus.

[0008] In some embodiments, the proximal 3' regulatory sequences are derived from a ferritin heavy chain locus, and in some embodiments the vector further includes distal 3' flanking sequences of a ferritin heavy chain locus.

[0009] In certain embodiments of the invention, the insertion site for a heterologous sequence includes at least one restriction endonuclease site, and in other embodiments the insertion site for a heterologous sequence is a polylinker site including at least two restriction endonuclease sites.

[0010] In certain embodiments of the invention, the proximal 5' regulatory sequences include a eukaryotic intron sequence. In some of these embodiments, the eukaryotic intron sequence is derived from intron 1 of a ferritin heavy chain gene. In certain embodiments, the proximal 5' regulatory sequences include untranslated exon sequences.

[0011] In some embodiments, the distal 5' flanking sequences and the proximal 5' regulatory sequences have a total length of between 1,000 and 10,000

bases. Similarly, in some embodiments, the proximal 3' regulatory sequences and any distal 3' flanking sequences have a total length of between 1,000 and 10,000 bases.

[0012] In another aspect, the invention provides eukaryotic cells transfected with any of the vectors of the invention. In some embodiments, the vector has stably
5 integrated into a chromosome of said cell and, in some embodiments, the first heterologous coding sequence is expressed in said cell.

[0013] In some embodiments, the invention provides eukaryotic cells including: (a) distal 5' flanking sequences of a eukaryotic locus; (b) proximal 5' regulatory sequences of a eukaryotic locus; (c) at least a first coding sequence; and
10 (d) proximal 3' regulatory sequences effective for transcription termination of a eukaryotic locus; in which the sequences are operably joined in order (a)-(d) in a 5' to 3' orientation, with optional linker sequences between adjacent sequences; and in which (1) the distal 5' flanking sequences comprise an exogenous sequence of at least 100 bases having at least 70% identity to a nucleotide sequence found between 20 bp and
15 100,000 bp 5' of a transcriptional initiation site of a ferritin heavy chain locus; and/or (2) the proximal 5' regulatory sequences comprise an exogenous sequence of at least 20 bases having at least 70% identity to a nucleotide sequence found between 1 bp and 10,000 bp 5' of a translational initiation codon of a ferritin heavy chain locus.

[0014] In another aspect, the invention provides a eukaryotic cell including an
20 exogenous 5' distal flanking sequence derived from a ferritin heavy chain locus operably joined to a coding sequence.

[0015] In another aspect, the invention provides a method of producing a desired protein in a eukaryotic cell including the steps of (a) providing at least one cell of the invention or a descendent thereof; (b) maintaining the cell in a culture under
25 conditions which permit high expression of the desired protein; and (c) isolating the desired protein from the culture.

[0016] These and other aspects and advantages of the invention will be apparent to those of skill in the art from the detailed description and examples which follow.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

5 [0018] Figure 1 shows rat ferritin heavy chain exon sequences.

[0019] Figure 2 illustrates one example of the subcloning of the region containing the ferritin heavy chain exons into the Litmus 38 plasmid.

[0020] Figure 3 illustrates the deletion of exons 2, 3, and 4 from pFerX1 and insertion of a polylinker to generate plasmid pFerX2.

10 [0021] Figure 4 illustrates the deletion of the exon 1 coding region from pFerX2 to generate plasmid pFerX3, and deletion of the IRE to generate plasmid pFerX4.

[0022] Figure 5 A-B illustrates the removal of exons 2 through 4 of the ferritin heavy chain gene from cosmid 15A using PCR fusion.

15 [0023] Figure 6 illustrates the insertion of the PCR fusion product of Figure 5 into the HpaI and AatII sites of pFerX4 to generate plasmid pFerX5.

[0024] Figure 7 illustrates the removal of the SmaI site from pFerX5 to generate plasmid pFerX5.1.

20 [0025] Figure 8 illustrates the addition of the distal 3' flanking sequences to pFerX6 to generate pFerX7.

[0026] Figure 9 illustrates the addition of the distal 5' flanking sequences of the ferritin heavy chain gene to pFerX7 to generate plasmid pFerX8.

[0027] Figure 10 illustrates the genetic map of plasmid pFerX8, including the sources of the sequences.

25 [0028] Figure 11 illustrates the genetic map of plasmid pFerX9, including the sources of the sequences.

[0029] Figure 12 illustrates the sequence of the transcribed region of the pFerX8 and pFerX9 plasmids.

30 [0030] Figure 13 illustrates the genetic map of pSIDHFR.2, a DHFR expression plasmid.

[0031] Figure 14 shows the results of experiments measuring reporter gene expression in pools of transfectants.

[0032] Figure 15 shows the results of experiments measuring reporter gene expression in transfected isolates.

DETAILED DESCRIPTION

5 [0033] The patent, scientific and medical publications referred to herein establish knowledge that was available to those of ordinary skill in the art at the time the invention was made. The entire disclosures of the issued U.S. patents, published and pending patent applications, and other references cited herein are hereby incorporated by reference.

10 Definitions.

 [0034] All technical and scientific terms used herein, unless otherwise defined below, are intended to have the same meaning as commonly understood by one of ordinary skill in the art; references to techniques employed herein are intended to refer to the techniques as commonly understood in the art, including variations on those
15 techniques or substitutions of equivalent techniques which would be apparent to one of skill in the art. In order to more clearly and concisely describe the subject matter which is the invention, the following definitions are provided for certain terms which are used in the specification and appended claims.

 [0035] Eukaryotic Locus. As used herein, the term "eukaryotic locus" refers
20 to any chromosomal genetic locus of a eukaryotic cell which encodes a polypeptide or RNA product which can be expressed in the cell under appropriate conditions. Mitochondrial loci are expressly excluded from the scope of the term "eukaryotic locus" as used herein.

 [0036] Distal 5' Flanking Sequences. As used herein, the term "distal 5'
25 flanking sequences" refers to flanking nucleotide sequences which are 5' of the proximal 5' regulatory sequences of a gene. Thus, although these sequences can have an effect on transcription rates because of their effects on chromatin structure, these sequences are generally 5' of the basic regulatory sequences (e.g., operators, promoters, ribosome-binding sites) and further removed from the transcriptional initiation site than
30 the proximal 5' regulatory sequences. The size of the distal 5' flanking sequences can range between 100-100,000 bases. In certain embodiments, the distal 5' flanking sequences will include between 500-50,000 bases, 750-25,000 bases or 1,000-10,000

bases. The distal 5' flanking sequences can begin anywhere 5' of the proximal 5' regulatory sequences, and typically begin 20 bases, 50 bases, 75 bases, 100 bases, 500 bases, 1,000 bases, 5,000 bases or 10,000 bases 5' of the transcription initiation site. Distal 5' flanking sequences can extend for substantial distances 5' of the promoter and transcriptional initiation sequences of a gene, and typically end 100,000 bases, 50,000 bases, 25,000 bases or 10,000 bases 5' of the transcription initiation site.

[0037] Proximal 5' Regulatory Sequences. As used herein, the term "proximal 5' regulatory sequences" refers to nucleotide sequences which are located near the 5' end of a gene and which include the basic regulatory elements (i.e., the promoter and, if present, operator and ribosome binding sequences) necessary for transcription and translation. The size of the proximal 5' regulatory sequences can range between 20-10,000 bases. In certain embodiments, the proximal 5' regulatory sequences will include between 50-5,000 bases, 75-1,000 bases or 100-500 bases. In some embodiments, the 3' end of the proximal 5' regulatory sequences can be defined as immediately 5' of the translation initiation or "start" codon of the coding region. Alternatively, in some embodiments, the proximal 5' regulatory sequences can include sequences internal to the gene including intron sequences and, therefore, the 3' end of the proximal 5' regulatory sequences can extend to the intron sequences. Moreover, in some embodiments, the proximal 5' regulatory sequences can include some 5' coding sequences (e.g., the start codon and/or a short N-terminal sequence). Proximal 5' regulatory sequences extend 5' of the transcriptional initiation site, and can end 10,000 bases, 5,000 bases, 1,000 bases, 500 bases, 100 bases, 75 bases, 50 bases or 20 bases 5' of the transcriptional initiation site.

[0038] Proximal 3' Regulatory Sequences. As used herein, the term "proximal 3' regulatory sequences" refers to nucleotide sequences which are located near the 3' end of a gene and which include the basic regulatory elements (i.e., the translational termination codon, polyadenylation signal and transcriptional terminator) necessary for proper mRNA processing and translation termination. The size of the proximal 3' regulatory sequences can range between 10-2,000 bases. In certain embodiments, the proximal 3' regulatory sequences will include between 25-1,000 bases, 50-750 bases or 75-500 bases. The 5' end of the proximal 3' regulatory sequences can be defined by the translational termination or "stop" codon (i.e., TAG, TTA or TGA). Proximal 3' regulatory sequences extend 3' of the translational termination codon, and can end

2,000 bases, 1,000 bases, 750 bases or 500 bases 3' of the translational termination codon.

[0039] Distal 3' Flanking Sequences. As used herein, the term "distal 3' flanking sequences" refers to flanking nucleotide sequences which are 3' of the proximal 3' regulatory sequences of a gene. Thus, these sequences are 3' of the basic regulatory sequences (i.e., the stop codon, and polyadenylation signal) necessary for proper mRNA processing and translation termination, and are further removed from the transcriptional termination site than the proximal 3' regulatory sequences. The size of the distal 3' flanking sequences can range between 100-100,000 bases. In certain embodiments, the distal 3' flanking sequences will include between 500-50,000 bases, 750-25,000 bases or 1,000-10,000 bases. The distal 3' flanking sequences can begin anywhere 3' of the proximal 3' regulatory sequences, and typically begin 500 bases, 750 bases, 1,000 bases or 2,000 bases 3' of the translation termination codon. Distal 3' flanking sequences can extend for substantial distances 3' of the transcriptional termination codon and polyadenylation sequences of a gene, and typically end 100,000 bases, 50,000 bases, 25,000 bases or 10,000 bases 3' of the transcriptional termination codon.

[0040] Vector. As used herein, the term "vector" means any genetic construct, such as a plasmid, phage, transposon, cosmid, chromosome, virus, virion, etc., which is capable transferring nucleic acids between cells. Vectors may be capable of one or more of replication, expression, recombination, insertion or integration, but need not possess each of these capabilities. Thus, the term includes cloning and expression vectors.

[0041] Transfection. As used herein, the term "transfection" means the introduction into a cell or an organism of a vector that replicates within that cell or organism or that expresses a polypeptide sequence in that cell or organism with or without integrating into the genome of that cell or organism. The term "transfection" is used to embrace all of the various methods of introducing such vectors, including, but not limited to the methods referred to in the art as transfection, transformation, transduction, or gene transfer, and including techniques such as microinjection, DEAE-dextran-mediated endocytosis, calcium phosphate coprecipitation, electroporation, liposome-mediated transfection, ballistic injection, viral-mediated transfection, and the

like. Cells or organisms which have undergone transfection are referred to herein as "transfectants."

[0042] Stable Transfection. As used herein, the term "stable transfection" means transfection, as defined above, which results in integration of all or a part of the vector into the genome of the transfected cell or organism. Cells or organisms which have undergone stable transfection are referred to herein as "stable transfectants."

[0043] Operably Joined. As used herein, the term "operably joined" refers to a covalent and functional linkage of genetic regulatory elements and a genetic coding region which can cause the coding region to be transcribed into mRNA by an RNA polymerase which can bind to one or more of the regulatory elements. Thus, a regulatory region, including regulatory elements, is operably joined to a coding region when RNA polymerase is capable under permissive conditions of binding to a promoter within the regulatory region and causing transcription of the coding region into mRNA. In this context, permissive conditions would include standard intracellular conditions for constitutive promoters, standard conditions and the absence of a repressor or the presence of an inducer for repressible/inducible promoters, and appropriate *in vitro* conditions, as known in the art, for *in vitro* transcription systems.

[0044] Heterologous. As used herein, the term "heterologous" means, with respect to two or more genetic sequences, that the genetic sequences are not operably joined in nature or do not naturally occur within the same genome in nature. For example, if a vector includes a coding region which is operably joined to one or more regulatory elements, these sequences are considered heterologous to each other if they are not operably joined in nature or they are not found in the same genome in nature.

[0045] Nucleotide Positions. As used herein, all nucleotide positions are designated with respect to the strand of DNA which includes elements of the ferritin heavy chain gene region in the "sense" orientation. As will be apparent from the context, numerical nucleotide positions are either designated with respect to the position of the start codon of the ferritin heavy chain gene or with respect to the position within one of the sequences included in the Sequence Listing. In the former case, the adenosine or "A" of the start codon (ATG) is designated as position 1, with preceding positions being negatively numbered. In the latter case, the relevant SEQ ID NO will always be specified. Relative nucleotide positions will be described with reference to the conventional 5' and 3' directions on the sense strand.

[0046] Percentages of Nucleotide Sequence Identity. As used herein, the percentage of sequence identity between two nucleotide sequences are calculated based upon the number of residues which are identical between the aligned sequences divided by the number of nucleotides present in the smaller of the two sequences. Before
5 calculation of the percentage identity, the sequences are aligned using the algorithm (or an equivalent algorithm) of the ClustalW program with default values, available through the European Bioinformatics Institute of the European Molecular Biology Laboratory (EMBL) (<http://www.ebi.ac.uk/clustalw>), and described in Higgins et al. (1994), "CLUSTAL W: Improving the sensitivity of progressive multiple sequence
10 alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Res.* 22:4673-4680.

[0047] Derived From. As used herein, the term "derived from," when used in relation to the origin of a nucleotide sequence, means that the sequence has been or can be obtained or produced, directly or indirectly, from a reference sequence by making a
15 limited number of insertions, deletions or substitutions in the reference sequence. Thus, for example, a sequence which is a subset of a reference sequence can be derived from the reference sequence by deleting flanking sequences. Similarly, a sequence can be derived from a reference sequence by a combination of insertions, deletions and/or substitutions of one or more nucleotides in a reference sequence. The number of
20 insertions, deletions and substitutions can be limited by a required percentage identity between the reference sequence and the derived sequence.

[0048] Numerical Ranges. As used herein, the recitation of a numerical range for a variable is intended to convey that the invention may be practiced with the variable equal to any of the values within that range. Thus, for a variable which is
25 inherently discrete, the variable can equal each integer value of the numerical range, including the end-points of the range. Similarly, for a variable which is inherently continuous, the variable can equal each real value of the numerical range, including the end-points of the range. As an example, a variable which is described as having values between 0 and 2, can be 0, 1 or 2 for variables which are inherently discrete, and can be
30 0.0, 0.1, 0.01, 0.001, or any other real value ≤ 2 for variables which are inherently continuous.

[0049] Or. As used herein, unless specifically indicated otherwise, the conjunction "or" is used in the "inclusive" sense of "and/or" and not the "exclusive" sense of "either/or."

General Considerations.

5 [0050] The present invention depends, in part, upon the development of a high expression "locus vector" derived from the ferritin heavy chain gene. The concept of a "locus vector" is based on the observation that the regions found 5' and 3' to highly expressed genes in their natural chromatin contexts can confer higher levels of expression to a heterologous gene. Therefore, the present invention provides a ferritin
10 heavy chain gene locus vector which includes 5' and 3' sequences which convey high levels of expression to heterologous genes in stable transfectants. Thus, the invention provides genetic vectors for the stable transfection and expression at high levels of a desired protein within eukaryotic cells.

The Ferritin Heavy Chain Gene.

15 [0051] The rat and human genomes contain multiple processed pseudogenes of the ferritin heavy chain (Hentze et al. (1986), *Proc. Natl. Acad. Sci. USA* 83:7226-72307). The rat ferritin gene consists of four exons (i.e., exons 1 through 4) separated by three introns (i.e., introns 1 through 3). GenBank Accession Nos. M18051, M18052 and M18053 disclose three gene segments which are shown in parts A, B, and C of
20 Figure 1. Together these three segments cover the four exons of the rat ferritin heavy chain genomic sequence. Figure 1(A) shows 168 bp of 5' untranslated sequence, including the transcriptional initiation site at position -168, followed by exon 1 and the first 104 bp of the 5' end of intron 1. Exon 1 includes the start codon and encodes 38 amino acids. Figure 1(B) shows the last 50 bp of the 3' end of intron 1, followed by
25 exon 2 and the first 35 bp of the 5' end of intron 2. Figure 1(C) shows the last 33 bp of the 3' end of intron 2, followed by exon 3, intron 3, exon 4 and 3' untranslated sequence, including the stop codon and polyadenylation signal 132 bp after the termination codon.

30 [0052] Because the insert sizes of cosmid libraries are quite large, they were chosen to obtain sufficient 5' and 3' flanking regions. In particular, rat cosmid library Catalog #RL1032m (BD Biosciences/Clontech, Palo Alto, CA) was selected. Other

libraries, however, also could have been used, or sequences could have been prepared synthetically.

[0053] In order to avoid cloning processed pseudogenes when screening the cosmid library, intron sequences were chosen to serve as probe templates. These
5 introns were cloned by PCR using rat genomic DNA (Catalog #6750-1, Clontech, Palo Alto, CA) as a template and primers based on related cDNA and genomic sequences from GenBank. Biotinylated probes were prepared using the introns as templates, and the cosmid library was screened with them. One ferritin heavy chain gene cosmid
10 (15A) was isolated and mapped with restriction enzymes. The three segments of rat genomic sequence from GenBank served as a guide to locate the coding regions and to plan the production of the high expression locus vector.

Production of Ferritin Heavy Chain Gene High Expression Locus Vector.

[0054] The production of a high expression locus vector of the invention can be accomplished in many ways. For example, the sequences forming the vector can be
15 obtained from a single clone or from multiple clones. The sequences can be based entirely on the rat ferritin heavy chain gene, entirely on another mammalian ferritin heavy chain, or on multiple mammalian ferritin heavy chain genes. The sequences can be based on all naturally-derived sequences or a mixture of naturally-derived and synthetic sequences. In addition, the locus vector can be produced by first obtaining
20 one or more large genomic fragments including all or part of the ferritin heavy chain gene region and then deleting or inactivating undesired sequences while inserting desired sequences, or can be produced by cloning or subcloning only the desired fragments of the ferritin heavy chain gene region and then combining these with other desired sequences. Similarly, mixtures of these approaches, employing cloning,
25 subcloning, deletion, inactivation and insertion can be employed to arrive at the desired construct. The approach taken and the order of the various steps is irrelevant to the invention and is within the discretion of one skilled in the art.

[0055] The high expression locus vectors of the invention include, in order from 5' to 3', (a) distal 5' flanking sequences of a eukaryotic locus; (b) proximal 5'
30 regulatory sequences of a eukaryotic locus; (c) at least a first insertion site for a heterologous sequence; and (d) proximal 3' regulatory sequences effective for transcription termination of a eukaryotic locus. Optionally, linker sequences may be

present between segments (a)-(d). Furthermore, at least one of the distal 5' flanking sequences and proximal 5' regulatory sequences has substantial identity with corresponding sequences of a ferritin heavy chain gene. In some embodiments, distal 3' flanking sequences are also included in the vector.

- 5 [0056] One embodiment of a high expression locus vector of the invention, the pFerX8 vector described below, is disclosed in GenBank Accession No. AY147930.

A. Distal 5' Flanking Sequences and Proximal 5' Regulatory Sequences.

- 10 [0057] In some embodiments, the distal 5' flanking sequences of the locus vector will include a sequence of 100-100,000 nucleotides having at least 70%-100% identity to a nucleotide sequence found within the distal 5' flanking sequences of a ferritin heavy chain locus. Thus, in some embodiments, the distal 5' flanking sequences can include at least 100, 500, 750, 1,000, 10,000, 25,000, 50,000 or 100,000 nucleotides having at least 70%, 75%, 80%, 85%, 90%, 95% or 100% identity to a
15 nucleotide sequence found within the distal 5' flanking sequences of a ferritin heavy chain locus. As shown in the examples below, the distal 5' sequences can include 1,000-10,000 bp, 2,000-9,000 bp, 3,000-8,000 bp or 4,000-7,000 bp of flanking sequences.

- 20 [0058] In other embodiments, the distal 5' flanking sequences of the locus vector will share lower percentages identity with the corresponding ferritin heavy chain gene sequences, and in some embodiments the distal 5' flanking sequences will be unrelated to any corresponding ferritin heavy chain gene sequences.

- 25 [0059] Downstream from the distal 5' flanking sequences, the high expression locus vector of the invention includes proximal 5' regulatory sequences. In some embodiments, the proximal 5' regulatory sequences of the locus vector will include a sequence of at least 20-10,000 nucleotides having at least 70%-100% identity to a nucleotide sequence found within the proximal 5' regulatory sequences of a ferritin heavy chain locus. Thus, in some embodiments, the proximal 5' regulatory sequences can include at least 20, 50, 75, 100, 500, 1,000, 5,000 or 10,000 nucleotides having at
30 least 70%, 75%, 80%, 85%, 90%, 95% or 100% identity to a nucleotide sequence found within the proximal 5' regulatory sequences of a ferritin heavy chain locus.

[0060] In other embodiments, the proximal 5' regulatory sequences of the locus vector will share lower percentages identity with the corresponding ferritin heavy chain gene sequences, and in some embodiments the proximal 5' regulatory sequences will be unrelated to any corresponding ferritin heavy chain gene sequences.

5 [0061] In all embodiments, the proximal 5' regulatory sequences must be effective to initiation transcription of the heterologous coding region to be inserted into the vector. Thus, in those embodiments in which the proximal 5' regulatory sequences are based upon the corresponding ferritin heavy chain gene sequences, they should not be varied to such an extent that the sequences become ineffective in initiating and
10 promoting transcription. Thus, the conservation of features such as the "TATA box" or ribosome binding site, or the replacement of these features with equivalent sequences, is necessary to preserve functionality of the expression vector. On the other hand, it is also acceptable to completely replace these sequences with functional equivalents from other genes, including any of the many known proximal 5' regulatory regions from
15 other genes. Similarly, it is acceptable to replace these sequences with chimeric sequences based upon the proximal 5' regulatory regions of two or more genes.

[0062] In some embodiments, both the distal 5' flanking sequences and the proximal 5' regulatory sequences include a sequence of at least 100-1000 nucleotides having at least 70%-100% identity to a nucleotide sequence found within, respectively,
20 the distal 5' flanking and proximal 5' regulatory sequences of a ferritin heavy chain locus. In some of these embodiments, the distal 5' flanking sequences and the proximal 5' regulatory sequences have 70-100% identity to contiguous sequences found within a ferritin heavy chain locus.

[0063] Because intron 1 of the ferritin heavy chain gene can contain positive
25 regulatory elements, and can aid in RNA processing and transport, it can be advantageous to create a locus vector that includes the maintenance of all or a portion of intron 1 as part of the proximal 5' regulatory sequences. This can be accomplished by maintaining an ATG codon and, optionally, additional codons 5' to the beginning of the intron 1 sequences. If codons other than the ATG are maintained, they can be
30 derived from the ferritin heavy chain gene exon 1 coding sequences or any other coding sequences (including synthetic or artificial sequences), and will encode the N-terminus of a fusion protein with the heterologous coding sequences. Such an N-terminus can function as a leader or signal sequence to aid in expression of the heterologous

sequences. Alternatively, in other embodiments, an additional heterologous sequence insertion site (e.g., a single restriction site or a polylinker) can be inserted 5' to the beginning of intron 1 so that sequences encoding various N-terminal sequences (e.g., leader or signal sequences) can be inserted at will. The ATG codon can be provided as part of the vector, or can be part of the inserted heterologous sequences.

[0064] However, there is no need to maintain either the ATG codon or any other codons prior to intron 1. Rather, in some embodiments, the ATG codon can be present in exon 2 or can be provided by a heterologous coding sequence. In such embodiments, the heterologous sequence insertion site will be present in exon 2, or at the intron 1/exon 2 junction, and the ATG codon either can be provided as part of the vector, or can be part of the inserted heterologous sequences. In all instances in which intron 1 is included in the vector, however, the splice donor and splice acceptor sequences of intron 1, or equivalent splice donor and acceptor sequences, must be maintained so that the intron sequences are post-transcriptionally removed. Other sequences within the intron can be deleted or varied, or additional sequences can be inserted, as described herein. However, in constructs in which intron 1 is maintained, insertion of a heterologous coding region, whether 5' or 3' of intron 1, must not disrupt the splice donor and acceptor sites, must reconstruct the splice donor and acceptor sites, or must provide equivalent splice donor and acceptor sites.

[0065] Finally, because the ferritin heavy chain gene exon 1 also contains an iron regulatory element (IRE) 3' to the ATG (at approximately positions -138 to -111) that negatively controls translation depending on the level of iron (reviewed in Klausner et al. (1993), *Cell* 72:19-26), the creation of the locus vector can optionally include the deletion of the IRE from the proximal 5' regulatory sequences.

B. Ferritin Heavy Chain Coding Regions.

[0066] Typically, the locus vector will not include any coding regions from the ferritin heavy chain gene. However, depending upon the method by which the vector is created, ferritin heavy chain coding regions can be included intentionally or as artifacts. For example, if the entire ferritin heavy chain gene region is cloned into a vector with the intention of using only the distal 5' flanking sequences and/or proximal 5' regulatory sequences (together "the 5' ferritin sequences"), the coding regions can be purposefully deleted in their entirety. Alternatively, a heterologous sequence insertion

site (e.g., a single restriction site or a polylinker) and proximal 3' regulatory sequences (and optionally distal 3' flanking sequences) could be inserted immediately 3' to the 5' ferritin sequences without deleting the coding regions. Because of the intervening insertion, the coding regions would be inactivated. Similarly, all of the coding regions except the start codon could be deleted or, alternatively, the heterologous sequence insertion site and proximal 3' regulatory sequences (and optionally distal 3' flanking sequences) could be inserted immediately 3' to the start codon. In addition, a larger portion of the coding region can be maintained before the insertion of the heterologous sequence insertion site and proximal 3' regulatory sequences (and optionally distal 3' flanking sequences) so that a fusion protein can be produced. Finally, combinations of the foregoing approaches can be employed such that the ferritin heavy chain coding regions are partially deleted and partially inactivated by the insertion of intervening sequences. In some embodiments, however, in order to reduce the size of the vector, inactivated and untranslated sequences are deleted.

15 C. Heterologous Sequence Insertion Site.

[0067] Downstream from the proximal 5' regulatory sequences, the high expression locus vector of the invention includes an insertion site for a heterologous sequence, such as a polylinker site. The heterologous sequence insertion site can be any sequence into which a heterologous sequence can be inserted in a sufficiently controlled and predictable manner to allow for production of functional high expression locus vectors with a reasonable expectation of success. Insertion sites for a heterologous sequence can include sites for homologous recombination, site-directed integration (e.g., via transposons or viral constructs), or endonuclease-mediated restriction. The length of the insertion site can vary from 4 bp (for use with four-cutter restriction endonucleases) to 1,000 bp or 5,000 bp (for use with homologous recombination methods). However, in certain circumstances, the 3' end of the proximal 5' regulatory sequences and the 5' end of the proximal 3' regulatory sequences can form an insertion site without the need for the inclusion of additional nucleotides between them. Thus, for example, the last two nucleotides of the proximal 5' regulatory sequences and the first two nucleotides of the proximal 3' regulatory sequences can form a 4 bp restriction site which can serve as an insertion site for the heterologous sequences. Alternatively, only one or a few nucleotides may be required to form an

insertion site between these sequences. Thus, the length of the insertion site could be 0, 1, 2, or 3 bp, as well as the 4 bp to 5,000 bp described above.

[0068] In some embodiments, the heterologous sequence insertion site will include one or more nucleotide sequences, on either the sense or antisense strand, which serve as restriction site(s) for natural or artificial endonucleases. These restriction sites can be unique in the vector, and the insertion site can be a polylinker that includes a multiplicity of such restriction sites to afford greater flexibility of use with different restriction endonucleases. An example of such a polylinker is provided in Example 1 and Figure 3.

10 D. Proximal 3' Regulatory Sequences.

[0069] Downstream from the insertion site for the heterologous sequences, the high expression locus vector of the invention includes proximal 3' regulatory sequences. At a minimum, these sequences include a polyadenylation signal. In some embodiments, the proximal 3' regulatory sequences also include a transcriptional termination signal. In some embodiments, the sequences can include the translation termination or stop codon, whereas in other embodiments the stop codon will be included in the heterologous sequence insert.

[0070] The proximal 3' regulatory sequences can be derived from the ferritin heavy chain gene, but need not be. For example, in some embodiments, the proximal 3' regulatory sequences of the locus vector will include a sequence of at least 10-2,000 bases nucleotides having at least 70%-100% identity to a nucleotide sequence found within the proximal 3' flanking sequences of a ferritin heavy chain locus. Thus, in some embodiments, the proximal 3' regulatory sequences can include at least 10, 25, 50, 100, 500, 750, 1,000, or 2,000 nucleotides having at least 70%, 75%, 80%, 85%, 90%, 95% or 100% identity to a nucleotide sequence found within the proximal 3' regulatory sequences of a ferritin heavy chain locus. In other embodiments, the proximal 3' regulatory sequences will consist essentially of a polyadenylation signal, which can be derived from a ferritin heavy chain gene, a heterologous sequence, or a synthetic or artificial sequence.

[0071] In other embodiments, the proximal 3' regulatory sequences of the locus vector will share lower percentages identity with the corresponding ferritin heavy

chain gene sequences, and in some embodiments the proximal 3' regulatory sequences will be unrelated to any corresponding ferritin heavy chain gene sequences.

E. Distal 3' Flanking Sequences.

[0072] Downstream from the proximal 3' regulatory sequences, the high expression locus vector of the invention optionally includes distal 3' flanking sequences. The distal 3' flanking sequences can be derived from the ferritin heavy chain gene, but need not be. For example, in some embodiments, the distal 3' flanking sequences of the locus vector will include a sequence of at least 100-100,000 nucleotides having at least 70%-100% identity to a nucleotide sequence found within the distal 3' flanking sequences of a ferritin heavy chain locus. Thus, in some embodiments, the distal 3' flanking sequences can include at least 100, 500, 750, 1,000, 10,000, 25,000, 50,000, or 100,000 nucleotides having at least 70%, 75%, 80%, 85%, 90%, 95% or 100% identity to a nucleotide sequence found within the distal 3' flanking sequences of a ferritin heavy chain locus. As shown in the examples below, the distal 3' flanking sequences can include 1,000-10,000 bp, 2,000-9,000 bp, 3,000-8,000 bp or 4,000-7,000 bp of flanking sequences.

[0073] In other embodiments, the distal 3' flanking sequences of the locus vector will share lower percentages identity with the corresponding ferritin heavy chain gene sequences, and in some embodiments the distal 3' flanking sequences will be unrelated to any corresponding ferritin heavy chain gene sequences.

[0074] The following examples illustrate some specific modes of practicing the present invention, but are not intended to limit the scope of the claimed invention. Alternative materials and methods may be utilized to obtain similar results.

25

EXAMPLE 1

Creation of a Ferritin Heavy Chain Locus Vector.

[0075] In order to generate a high expression locus vector based on the ferritin heavy chain gene, three phases of development were employed: (1) cloning of a ferritin heavy chain gene with substantial 5' and 3' regions; (2) production of an expression vector based on at least one of these gene regions, and (3) optimization of

the vector. As noted above, many other approaches could have been employed to produce the same or equivalent locus vectors.

[0076] First, the region containing the ferritin heavy chain exons from cosmid 15A was subcloned into the Litmus 38 vector (New England Biolabs) to generate plasmid pFerX1 (Figure 2). The BamHI-XhoI fragment was isolated from cosmid 15A and ligated into Litmus 38 digested with BamHI and SalI to generate plasmid pFerX1. Note that cosmid 15A was only partially sequenced and that some of the restriction site locations are based on restriction mapping. Therefore, some of the restriction site locations may not be accurate.

[0077] Figure 3 illustrates the deletion of the fragment containing exons 2, 3, and 4 from pFerX1 and the insertion of a polylinker containing AatII and SalI restriction sites to generate plasmid pFerX2. The deleted HpaI fragment extended from the HpaI site in the insert to the HpaI site in the vector in pFerX1. The 5' end of the polylinker regenerated the HpaI site, but the 3' end did not. Screening for the orientation of the linker was done using PCR.

[0078] The exon 1 coding region was deleted from pFerX2, leaving the ATG initiation codon and the following splice donor intact to generate plasmid pFerX3. Figure 4 illustrates that the deletion of the exon 1 coding region was accomplished by isolating the BamHI-BspHI (2515-2719) and NcoI-BamHI (2830-2515) fragments from pFerX2. BspHI and NcoI generate compatible overhangs which permitted the resulting fragments to be ligated together to generate pFerX3. As a result of this manipulation, exon 1 of the vector was changed from:

```

      BspHI
CCAGCCGCCATC ATG ACC ACC GCG TCT CCC TCG CAA GTG CGC CAG AAC TAC CAC CAG GAC TCG GAG GCT
GGTCGGCGGTAG TAC TGG TGG CGC AGA GGG AGC GTT CAC GCG GTC TTG ATG GTG GTC CTG AGC CTC CGA
      ▶Met Thr Thr Ala Ser Pro Ser Gln Val Arg Gln Asn Tyr His Gln Asp Ser Glu Ala

                                NcoI      Splice Donor
GCC ATC AAC CGC CAG ATC AAC CTG GAG TTG TAT GCC TCC TAC GTC TAT CTG TCC ATG GTGAGTGCAGCCT
CGG TAG TTG GCG GTC TAG TTG GAC CTC AAC ATA CGG AGG ATG CAG ATA GAC AGG TAC CACTCAGCCGGA
▶Ala Ile Asn Arg Gln Ile Asn Leu Glu Leu Tyr Ala Ser Tyr Val Tyr Leu Ser Met

```

to:

Splice Donor
 CCAGCCGCCATC ATG GTGAGTGC GGCCCT
 GGTCGGCGGTAG TAC CACTCACGCCGGA
 ▶ Met

[0079] Deletion of the exon 1 IRE was accomplished by replacing the SacII-EagI (2575-2639) fragment in pFerX3 with a linker that does not contain the IRE (but creates a 5' KpnI site for screening) to generate plasmid pFerX4. As a result of this
 5 manipulation, exon 1 of the vector was changed from (IRE underlined):

SacII (2575) EagI (2639)
 CAGAGTCGCCCGGGTT**TCCTGCTTCAACAGTGCTTGAACGGAA**CCCGGTGCTCGACCCCTCCGACCCCGTCCGGCCGCTTTGAGCC
 GTCTCAGCGCGCCAA**AGGACGAAGTTGTCACGAACTTGCTT**GGGCCACGAGCTGGGGAGGCTGGGGGCAGGCCGGCGAAACTCGG

to (linker shown in bold):

KpnI (2579) EagI (2611)
 SacII (2575)
 CAGAGTCGCCCGGGTACCGGTGCTCGACCCCTCCGACCCCGTCCGGCCGCTTTGAGCC
 GTCTCAGCGCGCCATGGCCACGAGCTGGGGAGGCTGGGGGCAGGCCGGCGAAACTCGG

[0080] A PCR fusion product was generated in a three step procedure to
 10 replace exons 2 though 4 with a polylinker containing SwaI and NotI, while
 maintaining the proximal 5' regulatory sequences and proximal 3' regulatory sequences
 of the ferritin heavy chain gene. As shown in Figure 5(A), the first PCR used cosmid
 15A (Figure 2) as a template. Primer locations for primers Fer1 and Fer4 are indicated
 by arrows. The "priming" region for primers FN1 and FN2 are also indicated by bars.
 15 In the second step, shown in Figure 5(B), a Fer1-FN2 PCR product was generated...
 The location of the "priming" region of primer Swa-2 is indicated. In the third step,
 shown in Figure 5(C), a FN1-Fer4 PCR product was generated. The location of the
 "priming" region of primer Swa-1 is indicated. In the fourth and final step, as shown in
 Figure 5(D), the final PCR fusion product was generated by using the Fer1-Swa-2 and
 20 Swa-1-Fer4 products as templates and the Fer1 and Fer4 primers. The HpaI-AatII
 fragment was isolated from this product for insertion into the HpaI and AatII sites of
 pFerX4 to generate plasmid pFerX5 (see Figure 6). The PCR fusion reactions used in
 the first three steps to generate the SwaI-NotI polylinker are shown in TABLE 1.

TABLE 1

| | Template(s) | 5' primer | 3' primer |
|------------|----------------------------------|--------------|-----------|
| First PCR | Cosmid 15A | Fer1 | FN2 |
| | Cosmid 15A | FN1 | Fer4 |
| Second PCR | Fer1/FN2 product | Fer1 | Swa-2 |
| | FN1/Fer4 product | Swa-1 | Fer4 |
| Third PCR | Fer1/Swa-2 & Swa-1/Fer4 products | Fer1 or Fer3 | Fer4 |

The PCR primers are shown below, where the polylinker sequence is shown in bold, and the complementary sequences between FN1 and FN2 or between Swa-1 and Swa-2 are shown underlined.

FN1 NheI NotI AatII
ACTTTCAGCTGCTAGCGGCCGCGCTGACGTCCCCAAGGCCAT

FN2 NotI NheI
ACGTCAGCGCGGCCGCTAGCAGCTGAAAGTGGAAAGGGTAT

Swa-1 SwaI NotI AatII
CTTTCATTAAATCTGCTAGCGGCCGCTGACGTC

Swa-2 SwaI
TAGCAGATTAAATGGAAAGGGTATTTGTTATTGATC

- 10 [0081] The SwaI site in the vector backbone of pFerX4 was removed by blunt cleavage of the plasmid with SwaI and insertion of the double-stranded oligo:

GGCGCGCC
CCGCGCGG

- 15 which contains an AscI site to generate plasmid pFerX4.1. The SwaI site was removed from the vector backbone in order to make the SwaI site in the polylinker above unique.

- [0082] The vector backbones of pFerX4.1 (in which the insertion of the AscI oligo of Figure 15 destroyed the SwaI site in the vector backbone) and pFerX5 (which
20 included the SwaI site in the backbone and the polylinker) were swapped using SacII-

AatII fragments to generate plasmid pFerX5.1 (Figure 7). pFerX5.1 contained the polylinker and but lacked the SwaI site in the backbone, making the SwaI site in the polylinker unique.

5 [0083] The polylinker

| | | |
|------------|-------------------------------|------------|
| | BglII | BstBI |
| | CTGTGAGATCTGTTTGAATGG | |
| | TGCAGACACTCTAGACAAGCTTACCAGCT | |
| AatII | | SalI |
| compatible | | compatible |

was inserted into the SalI-AatII sites of pFerX5.1 to generate plasmid pFerX6. The polylinker includes both BglII and BstBI sites and was designed to receive the distal 3' flanking sequences of the ferritin heavy chain gene.

10 [0084] The distal 3' flanking sequences of the ferritin heavy chain gene (AatII-BamHI fragment from cosmid 15A) were inserted into the AatII-BglII sites of pFerX6 to generate plasmid pFerX7 (Figure 8).

[0085] The distal 5' flanking sequences of the ferritin heavy chain gene (BamHI fragment from pFerH1, a subclone of cosmid 15A, Figure 2: BamHI 10269-15176) were inserted into the BamHI site of pFerX7 to generate plasmid pFerX8 (Figure 9).

[0086] The origins of the various sequences forming pFerX8 are shown in Figure 10. The Litmus 38 backbone is indicated by the filled box. This plasmid contains >6kb of distal 5' flanking sequences before the initiating ATG codon and ~7kb of distal 3' flanking sequences following the termination codon. The SwaI and NotI cloning sites are located at positions 10240 and 10254, respectively. Coding regions inserted into the SwaI and NotI sites should be blunt ended at the 5' end (SwaI end) and should start with the bases CAG to regenerate the splice acceptor followed by the second amino acid. The NotI site should be present at the 3' end following the termination codon.

[0087] An additional segment of distal 5' flanking sequence (BspEI fragment from cosmid 15A) was inserted into the BspEI site (6037) of pFerX8 to generate plasmid pFerX9 (Figure 11; BspEI fragment 6034-14211). This insertion adds both a unique segment of distal 5' flanking sequence as well as repeating a segment of the distal 5' flanking sequence already present in pFerX8 (10697-13990 is the same as

2520-5813). This plasmid is not entirely sequenced and the locations of some of the restriction sites are estimated based on restriction fragment sizes.

[0088] The sequence of the transcribed region of the pFerX8 and pFerX9 plasmids is shown in Figure 12. The putative transcription start site is indicated. The intron is shown in lower case. The putative TATA and polyadenylation signals are underlined. The initiation codon is in the first exon and the inserted gene, starting with the second amino acid, is inserted into the SmaI and NotI sites.

EXAMPLE 2

10 Expression of Heterologous Sequences.

A. Reporter Gene

[0089] A reporter gene was inserted into the SmaI-NotI sites in the polylinker of both the pFerX8 and pFerX9 plasmids. Secreted alkaline phosphatase (SEAP) was selected as a reporter gene because the commercially available assay (Clontech, Palo Alto, CA) for the product is simple and rapid. The expression vectors were designated pFerX8SEAP and pFerX9SEAP.

[0090] The sequence of the vector polylinker and the original sequence at the 5' end of exon 2 that needs to be recreated to regenerate the splice donor are shown in Figure 5. Thus, the 5' primer should include a CAG at the 5' end to recreate the natural splice donor followed by the coding region starting with the second amino acid (the ATG is already included in exon 1). The 5' end of the PCR product should be left blunt-ended for ligation with the SmaI site. For example:

General 5' primer:

25 CAG NNN NNN NNN NNN NNN NNN NNN
 AA2 AA3 AA4 AA5 AA6 AA7 AA8

Primer for SEAP example:

30 CAG CTG CTG CTG CTG CTG CTG CTG GGC

[0091] The 3' primer should include a NotI site followed by the 3' end of the gene including the termination codon (opposite strand). The PCR product should be digested with NotI to generate an end compatible with the NotI site in the polylinker. For example:

General 5' primer:

NNNN GCGGCCGC NNN NNN NNN NNN NNN NNN NNN NNN
 NotI 3' end of gene
 site

5

Primer for SEAP example (termination codon in bold):

TTTT GCGGCCGC AGC TCA TGT CTG CTC GAA GCG GCC

[0092] Ligation of the PCR product with the vector (digested with SmaI and
 10 NotI) does not recreate a SmaI site at the 5' end of the insert. Instead the ligated
 product contains a suitable splice acceptor at the "SmaI end." The inserted region will
 also contain the coding sequence from the second amino acid to the termination codon
 followed by the NotI site at the 3' end. For example:

15 After ligation generally:

CCATTT CAG NNN NNN NNN // NNN NNN NNN GCGGCCGC TGACGT

Example for SEAP:

CCATTT CAG CTG CTG CTG // CAG ACA TGA GCGGCCGC TGACGT

20

B. Transfections

[0093] The host used for transfections was the CHO DG44(E) cell line
 (Urlaub et al. (1986), *Somatic Cell Mol. Gen.* 12:555-566), which had been selected for
 growth and survival in serum-free media. This cell line was maintained in a spinner
 25 flask in serum-free media with added nucleosides. The cells used for transfection were
 in exponential growth. Either 2×10^6 or 5×10^6 cells were used for each transfection.

[0094] Reporter plasmids were co-transfected with a plasmid designated pSI-
 DHFR.2 encoding dihydrofolate reductase (DHFR) so that stable transfectants could be
 selected in the DHFR-host. The pSI-DHFR.2 plasmid includes a selectable marker and
 30 the *dhfr* gene driven by the SV40 promoter with the SV40 enhancer deleted (Figure
 13).

[0095] All DNA was prepared by Megaprep kit (Qiagen, Valencia CA). Prior
 to transfection DNA was EtOH precipitated, 70% EtOH washed, dried, resuspended in
 HEBS (20 mM Hepes pH 7.05, 137 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO_4 , 6 mM

dextrose), and quantitated prior to transfection. As a positive control, a plasmid which expresses SEAP with an SV40 early promoter/enhancer (pSEAP2, Clontech, Palo Alto, CA) was employed. Negative controls included an empty pUC18 vector (ATCC #37253, American Type Culture Collection, Manassas, VA) as a reporter control and a no DNA transfection as a transfection control.

[0096] Each transfection contained 50 µg of a reporter plasmid and 5 µg pSIDHFR.2. Equal plasmid weight was selected rather than equimolar amounts. From a molarity perspective there are differences on the order of 3-5 fold between the control reporters and the test reporters (TABLE 3). In each case the test reporter was lower than the control.

TABLE 3

| Reporter plasmid (50 µg each) | Plasmid size (kb) | Molar ratio to controls | DHFR (5 ug each) | Reporter gene |
|----------------------------------|----------------------|----------------------------|---------------------|------------------|
| pSEAP2 | 5.1 | 1 | pSIDHFR.2 | SEAP |
| pFerX8SEAP | 18.9 | 0.27 | pSIDHFR.2 | SEAP |
| pFerX9SEAP | 26.6 | 0.19 | pSIDHFR.2 | SEAP |
| pUC18 | 2.7 | | pSIDHFR.2 | none |
| No DNA | | | No DNA | none |

[0097] Cells and DNA were transfected by electroporation in 0.8 ml of HEBS using a 0.4 cm cuvette (BioRad, Hercules, CA) at 0.28 kV and 950 µF. After the electroporation pulse, the cells were allowed to incubate in the cuvette for 5-10 min at room temperature. They were then transferred to a centrifuge tube containing 10 ml of Alpha-MEM plus nucleosides (GIBCO, Gaithersburg, MD) with 10% dFBS (HyClone, Logan, UT) and pelleted at 1K rpm for 5 min. Resuspended pellets were seeded into T-flasks in Alpha-MEM without nucleosides with 10% dFBS and incubated at 36°C with 5% CO₂ in a humidified incubator until colonies formed.

[0098] TABLE 4 summaries seven experiments which were conducted. Transfections 1-3 were each performed in triplicate, and transfections 4-7 were performed once each.

TABLE 4

| Exp. # | Reporter plasmid (50 µg each) | DHFR (5 µg each) | Reporter gene |
|--------|-------------------------------|------------------|---------------|
| 1 | pSEAP2 | pSIDHFR.2 | SEAP |
| 2 | pFerX8SEAP | pSIDHFR.2 | SEAP |
| 3 | pFerX9SEAP | pSIDHFR.2 | SEAP |
| 4 | pFerX8 | pSIDHFR.2 | none |
| 5 | pFerX9 | pSIDHFR.2 | none |
| 6 | pUC18 | pSIDHFR.2 | none |
| 7 | No DNA | No DNA | none |

C. Transfection Efficiency

[0099] Approximately 2 weeks after the transfections, colonies had formed.

- 5 Stable transfectants were analyzed as either pools or isolates. Although all the pSI-DHFR.2-containing transfections produced colonies, the transfections containing the ferritin heavy chain locus vectors produced fewer colonies than did the controls. This was true whether or not the locus vector expressed a product. These results were surprising since the same amount of DNA was included in each transfection. Because
10 of the difference in transfection efficiency it is recommended that multiple transfections be done to account for the reduced number of transfectants.

D. SEAP Assay

- [0100] The reporter constructs containing the SEAP gene were analyzed using the Great EscAPE™ SEAP Reporter System 3 (Clontech, Palo Alto, CA). This assay
15 uses a fluorescent substrate to detect the SEAP activity in the conditioned media. The kit was used in a 96-well format according to the manufacturer's instructions with the following exceptions. All standards and samples were diluted in fresh media rather than the dilution buffer provided. Instead of performing one reading after 60 min, multiple reads were taken at 10-20 min intervals and used to express SEAP activity as
20 relative fluorescent units per minute (RFU/min). The emission filter available for the Cytofluor II plate reader was 460 nm instead of the recommended 449 nm.

- [0101] All of the data generated for the pools and isolates below was based on the reporter constructs expressing SEAP. The titers reported were based on a positive control with the kit. Although absolute values were not derived, the relative titer values
25 are useful.

- [0102] The specific productivity was assessed in assays in which the media was exchanged for fresh media and then, 24 hours later, the media was sampled and the

cells were counted. The product titer was normalized for the cell number at the end of the 24 hour assay. Because the titers were relative, the specific productivities are expressed as relative values.

5 Specific productivity = $\frac{\text{product titer (/ml)} \times \text{volume (ml)}}{\text{time (days)} \times \# \text{ of cells}}$

E. Transfectant Pools

[0103] After the appearance of colonies, the cells were collected and pooled from each transfection. Pools were seeded into 6-well plates or T-flasks and were kept subconfluent for the 24 hour assay. Results from the pool assays are shown in Figure 14. Five pools were analyzed for each construct, two from experiment 1 (1A and 1B) and three from experiment 2 (2A, 2B, and 2C). All assays were done three to four weeks post-transfection. Note that the experiment 2C with pFeiX8SEAP had a very low transfection efficiency relative to the other transfections.

[0104] Specific productivities were fairly consistent with the control (pSEAP2) but highly variable with the pFerX8SEAP and pFerX9SEAP vectors. Notably, the ferritin vectors were capable of generating pools with higher specific productivities than the control.

20 F. Transfectant Isolates

[0105] Isolates were obtained by “picking” colonies from transfection experiment #2. “Picking” was accomplished by aspirating directly over a colony with a P200 Pipetman set at 50 μ l. The aspirated colony was transferred first to a 48-well plate and then to a 6 well plate when there were a sufficient number of cells. Specific productivities were assessed in 6-well plates at near confluent to confluent cell densities using the 24-hour assay described above. 40-50 isolates were analyzed for each construct. The results are shown in Figure 15, in which the isolates are presented in the order of their specific productivity for each SEAP expression construct. The scale of specific productivity is consistent between the panels for comparison.

30 **[0106]** The majority of the isolates (63%) from the pSEAP2 transfections did not express product above the limit of detection. The highest productivity from pSEAP2 in this experiment was 46 units per cell per day (relative value for comparison). In contrast only 28% of the isolates from the pFerX8SEAP transfections

expressed product below the limit of detection and 44% had productivities above the highest pSEAP2 transfectant. The highest productivity from pFerX8SEAP in this experiment was 259 units per cell per day, more than five-fold higher than the highest productivity from pSEAP2. Although the pFerX9SEAP construct performed better than pSEAP2, it did not perform as well as pFerX8SEAP.

EXAMPLE 3

Reduction of Vector Size.

[0107] In order to reduce the size of the vector for ease of use, 5' and/or 3' regions of the vector were deleted (TABLE 5). These deletions were tested as before using SEAP as a reporter. Approximately 30 isolates were tested from each of the plasmids shown in TABLE 5 as well as from the controls, pSEAP2 and pUC18 (10 isolates).

TABLE 5

| Plasmid | Region deleted | 5' end of the deletion* | 3' end of the deletion* | Size of the plasmid (bp)** |
|-------------|----------------|-------------------------|-------------------------|----------------------------|
| pFerX8SEAP | none | | | 19340 |
| pFerX10SEAP | 5' | 2513 | 7414 | 14439 |
| pFerX11SEAP | 3' | 13727 | 17636 | 15431 |
| pFerX12SEAP | 5' | 2513 | 7414 | 8042 |
| | 3' | 12704 | 19101 | |

* The deletion end points are based on the pFerX8 sequence numbering

** The SEAP gene constitutes 1557 bp of the plasmid

[0108] The pFerX11SEAP vector performed similarly to the pFerX8SEAP vector, indicating that the ~3.9 kb deletion in the 3' region described in TABLE 5 was not detrimental. The pFerX10SEAP and pFerX12SEAP vectors did not perform as well as pFerX8SEAP, indicating that the ~4.9 kb 5' deletion described in TABLE 5 was detrimental to function.

EQUIVALENTS

[0109] While this invention has been particularly shown and described with references to certain specific embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims. Those skilled in the art will recognize, or be able to ascertain using no more than

routine experimentation, many equivalents to the specific embodiments of the invention described specifically herein. Such equivalents are intended to be encompassed in the scope of the appended claims.

CLAIMS

What is claimed is:

1. A genetic vector for stable transfection and expression of a desired protein
5 within eukaryotic cells comprising:
 - (a) distal 5' flanking sequences of a eukaryotic locus;
 - (b) proximal 5' regulatory sequences of a eukaryotic locus;
 - (c) at least a first insertion site for a first heterologous coding sequence; and
 - (d) proximal 3' regulatory sequences effective for transcription termination of a
10 eukaryotic locus;wherein said sequences are operably joined in order (a)-(d) in a 5' to 3' orientation, with optional linker sequences between adjacent sequences; and
wherein
 - (1) said distal 5' flanking sequences comprise a sequence of at least 100 bases
15 having at least 70% identity to a nucleotide sequence found between 20 bp and 100,000 bp 5' of a transcriptional initiation site of a ferritin heavy chain locus; or
 - (2) said proximal 5' regulatory sequences comprise a sequence of at least 20 bases having at least 70% identity to a nucleotide sequence found between 1 bp and 10,000 bp 5' of a translational initiation codon of a ferritin heavy chain locus.
20
2. A genetic vector for stable transfection and expression of a desired protein
within eukaryotic cells comprising:
 - (a) distal 5' flanking sequences of a eukaryotic locus;
 - (b) proximal 5' regulatory sequences of a eukaryotic locus;
 - 25 (c) at least a first heterologous coding sequence encoding said desired protein;
and
 - (d) proximal 3' regulatory sequences effective for transcription termination of a eukaryotic locus;wherein said sequences are operably joined in order (a)-(d) in a 5' to 3'
30 orientation, with optional linker sequences between adjacent sequences; and
wherein

(1) said distal 5' flanking sequences comprise a sequence of at least 100 bases having at least 70% identity to a nucleotide sequence found between 20 bp and 100,000 bp 5' of a transcriptional initiation site of a ferritin heavy chain locus; or

5 (2) said proximal 5' regulatory sequences comprise a sequence of at least 20 bases having at least 70% identity to a nucleotide sequence found between 1 bp and 10,000 bp 5' of a translational initiation codon of a ferritin heavy chain locus.

3. A genetic vector as in any one of claims 1-2 wherein said distal 5' flanking sequences are derived from a ferritin heavy chain locus.

10

4. A genetic vector as in any one of claims 1-2 wherein said proximal 5' regulatory sequences are derived from a ferritin heavy chain locus.

5. A genetic vector as in any one of claims 1-2 wherein said proximal 5' regulatory sequences and said distal 5' flanking sequences are derived from a ferritin heavy chain locus.

15

6. A genetic vector as in any one of claims 1-5 wherein said proximal 3' regulatory sequences are derived from a ferritin heavy chain locus.

20

7. A genetic vector as in any one of claims 1-6 further comprising distal 3' flanking sequences of a ferritin heavy chain locus.

8. A genetic vector as in any one of claims 1, and 3-7 wherein said insertion site for a heterologous sequence includes at least one restriction endonuclease site.

25

9. A genetic vector as in claim 8 wherein said insertion site for a heterologous sequence is a polylinker site including at least two restriction endonuclease sites.

30 10. A genetic vector as in any one of claims 1-9 wherein said proximal 5' regulatory sequences include a eukaryotic intron sequence.

11. A genetic vector as in claim 10 wherein said eukaryotic intron sequence is derived from intron 1 of a ferritin heavy chain gene.
12. A genetic vector as in any one of claims 1-11 wherein said proximal 5' regulatory sequences include untranslated exon sequences.
13. A genetic vector as in any one of claims 1-12 wherein said distal 5' flanking sequences and said proximal 5' regulatory sequences have a total length of between 1,000 and 10,000 bases.
14. A genetic vector as in any one of claims 1-12 wherein said proximal 3' regulatory sequences and any distal 3' flanking sequences have a total length of between 1,000 and 10,000 bases.
15. A eukaryotic cell transfected with a vector of any one of claims 1-14.
16. A eukaryotic cell as in claim 15 wherein said vector has stably integrated into a chromosome of said cell.
17. A eukaryotic cell as in any one of claims 15-16 wherein said first coding sequence is expressed in said cell.
18. A eukaryotic cell comprising
- (a) distal 5' flanking sequences of a eukaryotic locus;
 - (b) proximal 5' regulatory sequences of a eukaryotic locus;
 - (c) at least a first coding sequence; and
 - (d) proximal 3' regulatory sequences effective for transcription termination of a eukaryotic locus;
- wherein said sequences are operably joined in order (a)-(d) in a 5' to 3' orientation, with optional linker sequences between adjacent sequences; and wherein

- (1) said distal 5' flanking sequences comprise an exogenous sequence of at least 100 bases having at least 70% identity to a nucleotide sequence found between 20 bp and 100,000 bp 5' of a transcriptional initiation site of a ferritin heavy chain locus; or
- (2) said proximal 5' regulatory sequences comprise an exogenous sequence of at least 20 bases having at least 70% identity to a nucleotide sequence found between 1 bp and 10,000 bp 5' of a translational initiation codon of a ferritin heavy chain locus.
19. A eukaryotic cell comprising:
an exogenous 5' distal flanking sequence derived from a ferritin heavy chain locus operably joined to a coding sequence.
20. A method of producing a desired protein in a eukaryotic cell comprising:
(a) providing at least one cell of any one of claims 15-19 or a descendent thereof;
(b) maintaining said cell in a culture under conditions which permit high expression of said desired protein; and
(c) isolating said desired protein from said culture.

A. Exon 1

BseI BseRI SfiI EarI
 AGCTCAGAGACCCAGAGCCGCTCACAATCACACAGGCTCTCCCGCCACGCACTGCTGGCTTGGGCAACACGCTACAGGAAGAGGCGGGCTGGGCGGCCACCGCGCT
 TCGAGTCTCTGGGTTCTCGGCGGAGTGTATGTGTCCGAGGAGGGCGGGTGCCTGACGACCGAACCCGTTGTGCGGATGTCTCTCCGCCCGACCGCGGGTGGCGGA
 EaeI BssHII EcoNI BamHI SmaI
 GATTGGCCGGAGCGCGCTGACGCGAGGATCCCGCTATAAAGTCGCGCCCGCTGGTCCCTACGCCAGAGCTTCTCGCCAGAGTTCGCGCGGTTTCTGCTTCAACAGTGTGGA
 CTAACCGGCTTCGCGCGGAGTGCCTCTAGGCGATATTCACGCCCGGGCGACAGGATGCGGTCTGCAAGAGCGGGTCTCAGCGCGCCCAAGGACGAAATTGTACGAACT
 BshIKAI PstAI EagI Bpu10I BsrBI
 ACGGAACCGGCTGTCTGACCCCTCCGACCCCGTCCGSCCGCTTTGAGCCTGAGCCCTTTGCAACTTCGTGCTCCGCTCCGCTCCAGCGTCCGCTCCGCGCCCTCGTCCAGCCGCG
 TGCCTTGGGCCACGAGCTGGGAGGCTGGGGGCGAGGCCGCGAAACTCGGACTCGGGAAACGTTGAAGCAGCGAGGCGCGAGGTCCGACGCGAGGCGCGAGGAGGTCCGCGG
 DrrI
 BspHI BsmBI BpmI NcoI
 ATCATGACCAACCGCTCTCCCTCGCAAGTGGCGCAGAACTACCAACAGGACTCGAGGGCTGCCATCAACCGCCAGATCAACCTGGAGTTGTATGCTCTACGTCTATCTGTCC
 TAGTACTGTGGCGCAGAGGGAGCGTTCACGCGGTCTTGATGCTGCTCTGAGCCTCCGACGCTAGTTGGCGGTCTAGTTGGACCTCAACATACGAGGATGCAGATAGACAGG
 Met Thr Thr Ala Ser Pro Ser Gl nVal Arg Gl nAsn Tyr Hi s Gl nAsp Ser Gl uAl aAla l eAsn Arg Gl n l eAsn Leu Gl uLeu Tyr Ala Ser Tyr Val Tyr Leu Ser
 DralI
 ATGGTGAAGTCGCGCTGGCTTTGCGGGGCGGAAAGAGGGTGGCGCTGGCTCCCTTGGGCCACTTGGTGAAGTGGCGAGGGTGGGTTGGGCGTGGCTCGCGG
 TACCACTCACGCCCGACCGGAAACGCCCGCGCTTCTCCACGCCCGGACCGAGGGGAACCGGTGAACCACTCGACCGCCTCCCAACCAACCCCGCACCGAGCGCC
 Met

B. Exon 2

Eco57I BspI
 GCATCTGCCTGCTGTGGGATCAATAACAAATACCCCTTCCACTTTCACTCTTGTATTTTGACCGGATGATGTGGCCCTGAAGAACTTTGCCAAATACTTTCCATCAATCT
 CTAGACGGAACGACACCCCTAGTTATTTATTTATGGGAAAGGTGAAAGTACAGAACATATAAATCTGGCCCTACTACACGGGACTTTTGAAACGTTTATGAAAGAGGTAGTTAGA
 Ser Cys Tyr PheAsp ArgAspValAlaLeuLysAsnPheAlaLysTyrPheLeuHIsGl nSer
 EarI NspI PstI BsdI SbfI EcoRV AccI
 CATGAAGAGAGGGAACATGCTGAGAACTGATGAAGCTGCAGAACGAGCGAGGTGGACGAATCTTCTGCGAGGATCAAGGTAAGTAGACTATGGGACTGCGTTAAATGAGCAGT
 GTACTTCTCTCCCTTGTACGACTCTTTGACTACTTCGACGCTCTGGTCCCTCCACCTGCTTAGAAGGACGCTCTATAGTTCCATTCATCTGATACCCCTGACCAATTTACTGTCA
 HIsGl uSer Gl uAl aAla l eLysLeuMetLysLeuGl nAsn Gl nArg Gl yAla l ePheLeuGl nAsp l eLys

C. Exons 3 and 4

PstI AflIII BmrI BsrBI BsmI BsrDI ApaLI BspI
 CTGCAGATGAATGACATGTTTCTTTGATTGAGAACTGACCGTGATGACTGGGAGAGCGGGCTGAATGCAATGAGGTGTGCACTGCACTTGGAAAGAGGTGATGATCAGTCA
 GACGCTACTTAACTGTACAAAGAACTAAATCTTTGACTGGCACTACTGACCCCTCTCGCCGACTTACGTTACTCCACAGTGTACGTGAACCTTTTCTCACACTTACTGAGT
 Lys Pro Asn ArgAspAspTrpGl uSer Gl uLeuAsnAlaMetArgCysAlaLeuHisLeuGl uLysSerValAsnGl nSer
 Pml DralI
 CTACTGGAACTTCAAACTTGGCTACTGACAAGATGATCCCACTGAGTATCAGAAACACGGGTGAGTGGAGATGATTGGCCACAGGGCTTGGGAGAGCTGACCACTAACC
 GATGACCTTGAAGTGTGGACCGATGACTGTTCTTACTAGGGGTGCACTCATAGTCTTTGTGCCCACTCACCTCTACTAAACGGTGTCCCGAACCTCTCGACTGGTCACTGG
 LeuLeuGl uLeuHisLysLeuAlaThrAspLysAsnAspProHis
 BsmBI BspMI XcmI BmrI BstEI PmlI
 CTGTCCCATGTTCTCTTCTAGTTATGTGACTTCATTGAGACGCACTTACTGAAATGAGCGAGGTGAATCCATTAAAGAACTGGGTGACCACTGACCACTTACCGAAGATGG
 GACAGGGTACAGAGAAAGATCAATACACTGAAGTAACTCTGCTAATGGACTTACTCTGCCACTTATAGTAAATTTCTTGAACCACTGGTGCACCTGGTGAATGCGTTCTACC
 LeuCysAspPhe l eGl uThrHisTyrLeuAsnGl uGl nValLysSer l eLysGl uLeuGl uAspHisValThrAsnLeuArgLysMetG
 BanII MstII BspI AatII StyI BstAI
 GAGCCCTGAAATCTGGCATGGCAGAAATCTCTTTGACAGCACACCTGGGACACGGTGATGAGAGCTAAGCTGACGTCCCAAGGCCATGTGACTTTACTGGCTCACTGAGG
 CTCGGGACTTAGACCGTACCGCTCTTATAGAGAACTGTTCTGTGGGACCTGTGCCACTACTCTCGATTGACTGCGAGGGTCCGGTACACTGAAATGACCGAGTGAAGTCC
 l yAlaProGl uSer Gl yMetAlaGl uTyrLeuPheAspLysHisThrLeuGl yHisGl uAspGl uSer***
 Ppu10I NsiI EcoT22I XapI AclI KpnI BstAI
 SphI
 CAGTGCATGCTAGTCCAGGCTGCTTTATCTTTCTATAAGTTGACCAAAACATCTGCTTAAAGTTCTTTAATTTGTACCAATTTCTTCAAATAAAGAAATTTGGTACCCAGCT
 GTCAAGTACGTACAGTCCGACGGAATAGAAAAGATATCAACGTGGTTTTGTAGACGAATTTCAAGAAATTAACATGGTAAAGAAATTTATTTCTTAAACCATGGGTGCA
 SspI
 CTGTGTGATTGAGGATGAGCGCACCGACTTCCCTTGGCTCGGCTATATAACCACTGCAACCGCTGAAAGAAATATTTATTAACCTGTAAGTTGGGAAAGATAGTGAAGA
 GAACACACTAATCTCTACTCGCTGGTGAAGGGAACCGACCGGATATATGGTGTGACGTGCGGACTTCTTATAAATAATTTGAGCATCAACCCCTTCTATCACTTTCT
 DseDI StyI NcoI XmnI SmlI
 CAGGTGTGTTGACAGGACTAAGCAGTCTGCTGCTGACCTGACCTGACGACTGCCATGGGACATATTCTTGAGTGTG
 GTCCACACAGTCTGTCTGATTCTGTCAGGACCAAGACTCAATGGACGCTGACGCTACCCCTTGTATAAGAACTCACAG

FIGURE 1

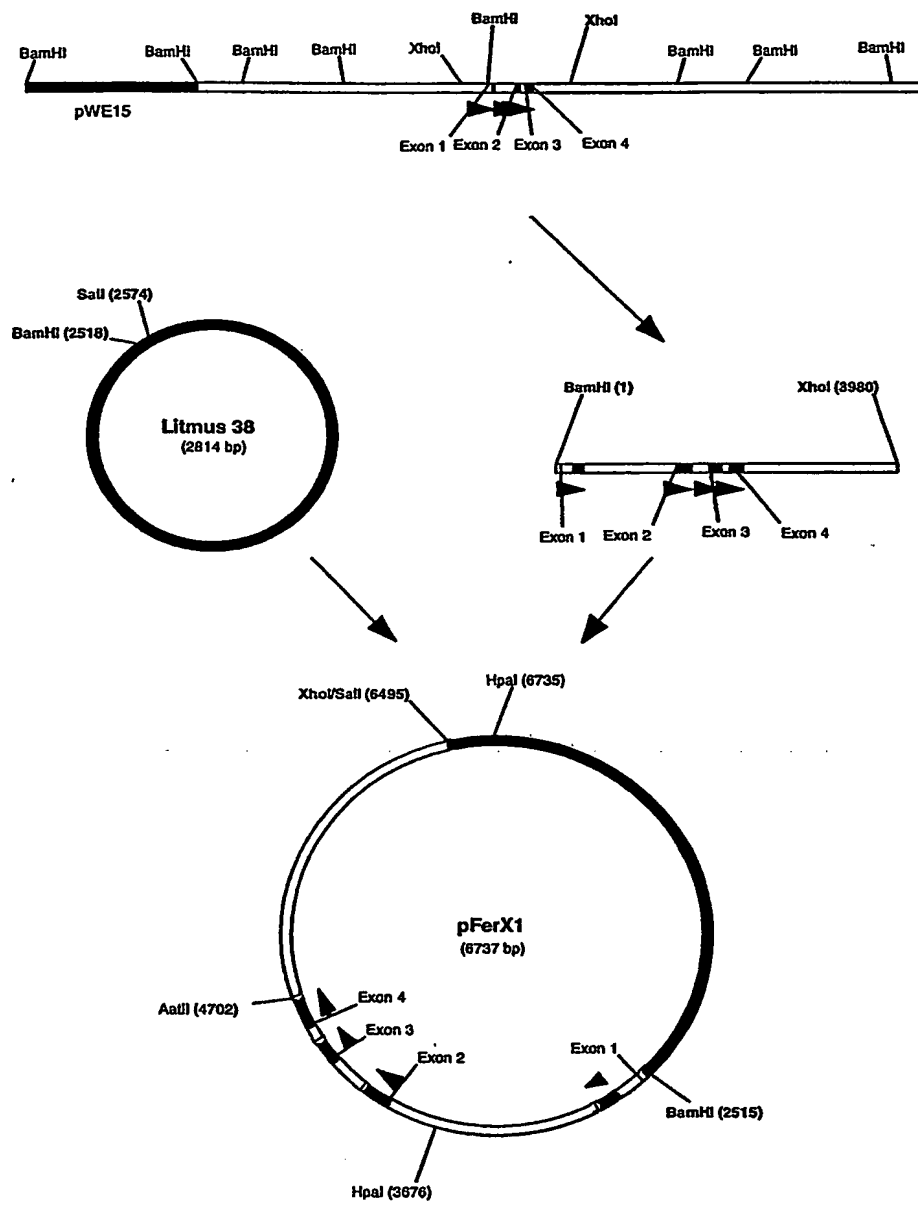


FIGURE 2

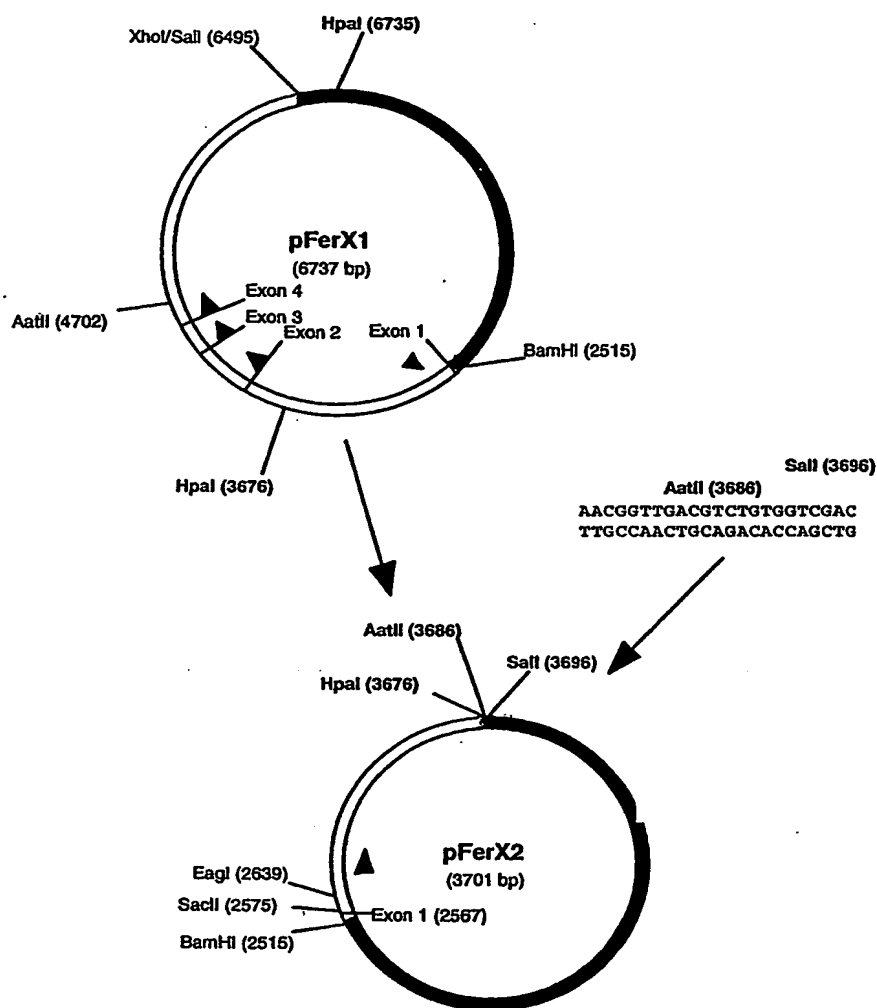


FIGURE 3

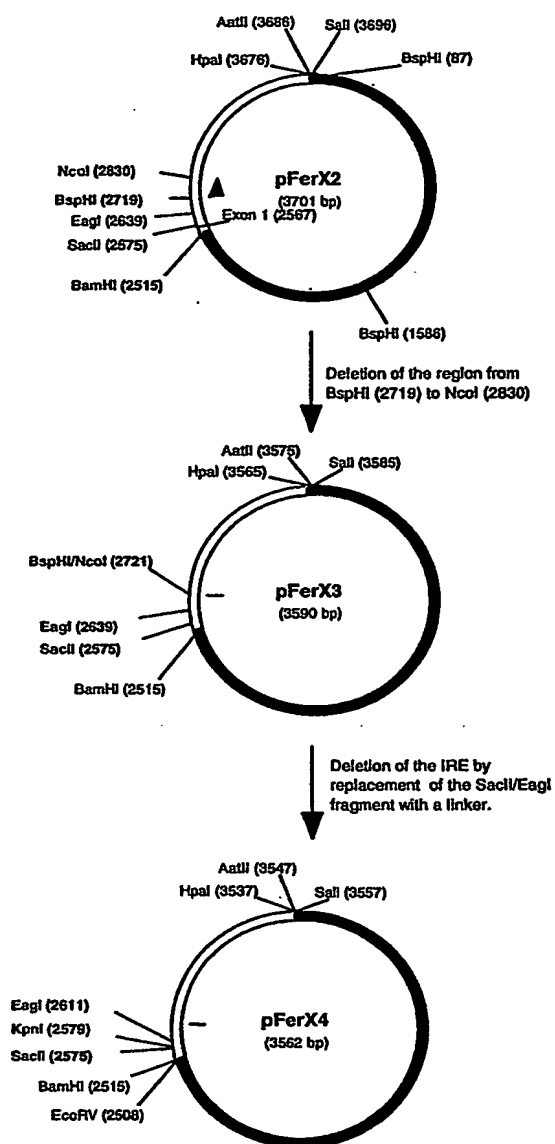


FIGURE 4

A.

[illegible]

B.

Fer1
ATC TGT CCA TG GTGAGTCGGCCTGGCCCTTTGGCGGGCGGAAAGAGGGTGCGGCCCTGGCCCTCCCTTGGGCCACTTGGTGAGCTGGCGGAGGG

↓

TGGGTGGGGCGTGGCCTGCTGCGGGCTTCCCGCCTTCCAGCGCCCTTCTGGAAATGGAGTTTGTCCGGGGTCTTTCCAAAGGCAGGCAGCCCTGCCGTGGCAAGTCTGAGCAGCTAGCGCTTTGTGGCTCCTGCATAGACAGGCACGTCATAACACCCGTTTGTGAAGCCTTAGGGCTGTACAACCTGTCAAGCTCTCCAATCAACCTGCAGTTAGGTGCATTTCCTGCAGCTCTGCTCCCTCCGGTCACCTGGCCTGCAGGCTCTCTGTGTGGGTGTACATCAGCTCCAGTCTCTGACTATGCGGGTCTGCTGTGTGATGGTGGAAATGAGCAGCCTGGGGCTTGGTATCAAAAGAGGCTATCTCTTGTGAACTTACTCTAACCACTTCTGAAGCAGCGCCTCTACATCTCTGCTTATCACAGAGCCTCACTTGCATTGAAACTTATCGCTAGGAATCTCCCTTCTGTAA TCACCCTGACCTTGCCCAAGGCATCTAGAGTACTGTACGTTTTTAATTTTTATTTTGCACCAAGTTGTGTGCTTACTAACAGAAGTAGTAGGTAACATAC TTGTTGGAAAAAGCCCGCTTGGGAAAAAACCAATTATCTGTGGAATACAAATACACTGAGTGCTCAAGCTAAAATCGAATAATCAAGCTCTCCAATGTAT

HpaI

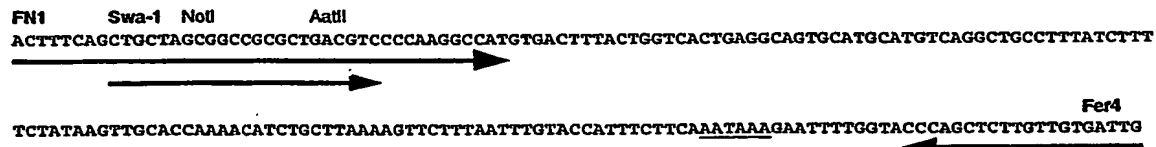
TTGTGCTAAAATACAATGCGCCTCAGTTCTTAAACCAAGGTAATCAGCAGTTGGCTGTCTAGCTGAAAACCTTGAGACCTTGTGTCTTAACCATTTTTTTA TTTACATGATTGTGAAGGAGAGAAATTGACCTCCCAATGTAGGGCAGTTTAGCACCCCCCTCTCAGACAAATAGATATGGCCTTGGCTTAAAGTT TTTTCTCTGCACATAATGTGGAGCCATAGAACCCTTGATAAAGCCAAAGTCCCAAGTTTGTTTTCCATCTTACTTTAAAGGCCAAGTAGGGTGACAA

SwA-2 NotI FN2

ACAGCCTTTACCACCATTTGCATCTGCCTTGCTGTGGGGATCAATAACAAATACCCCTTTCCACTTTCAGTCTGTAGCGGCCGCTGACGT

FIGURE 5A

C.



D.

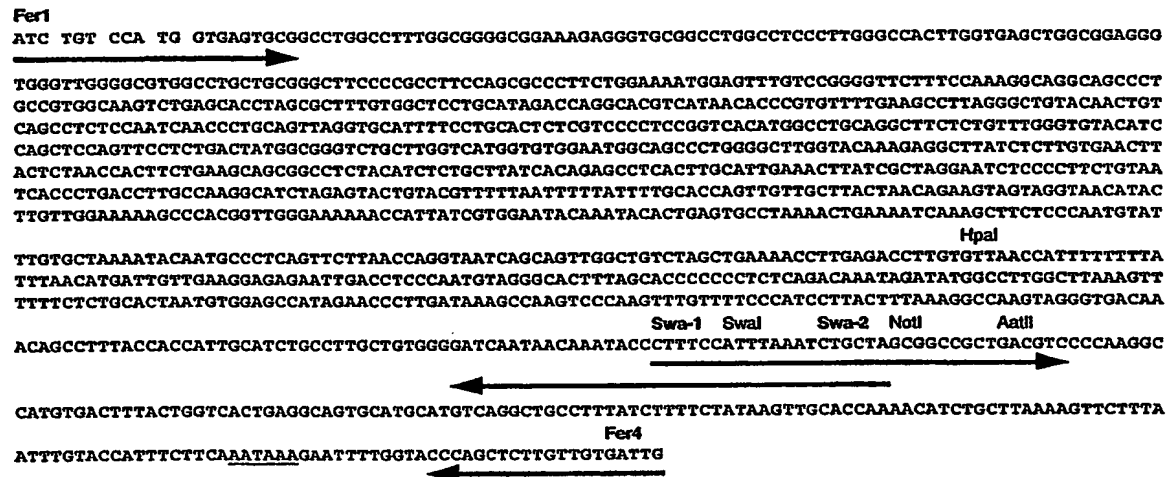


FIGURE 5B

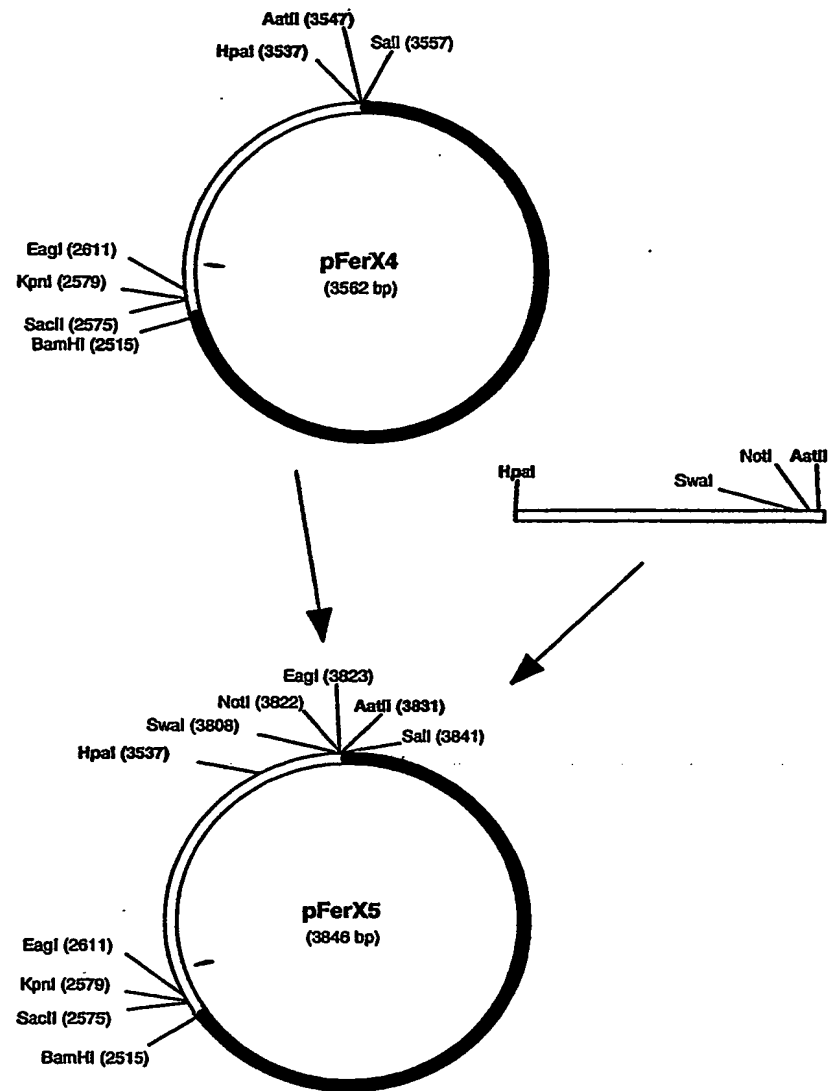


FIGURE 6

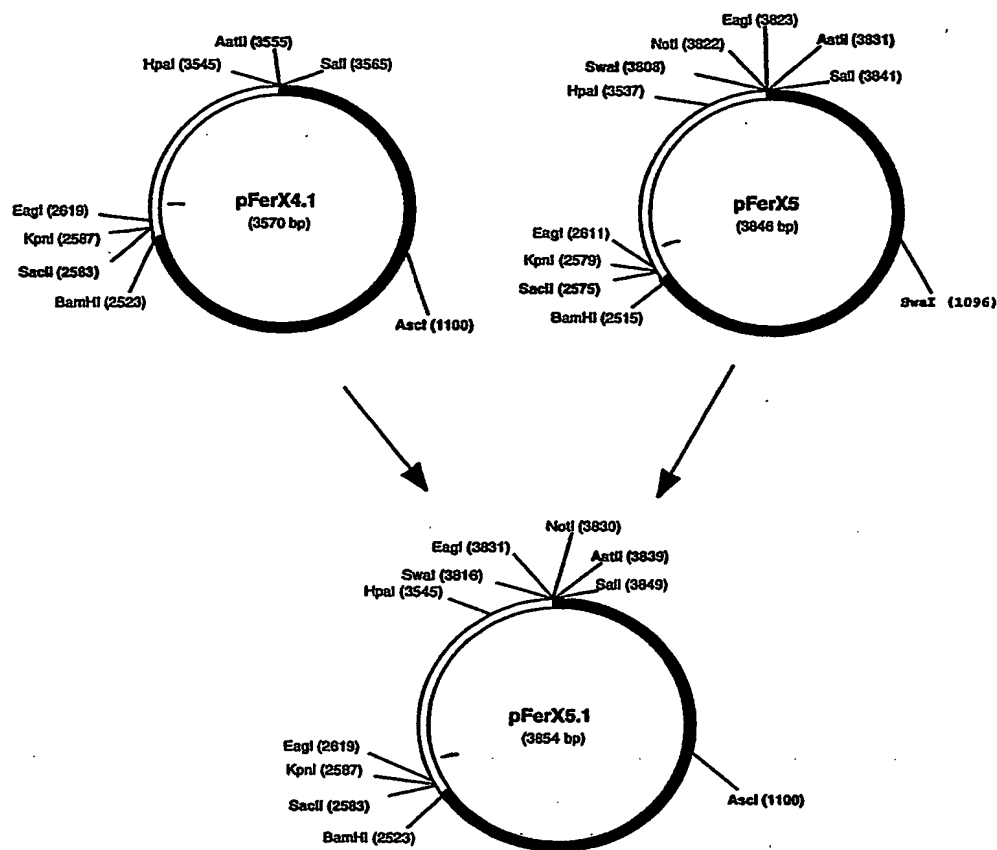


FIGURE 7

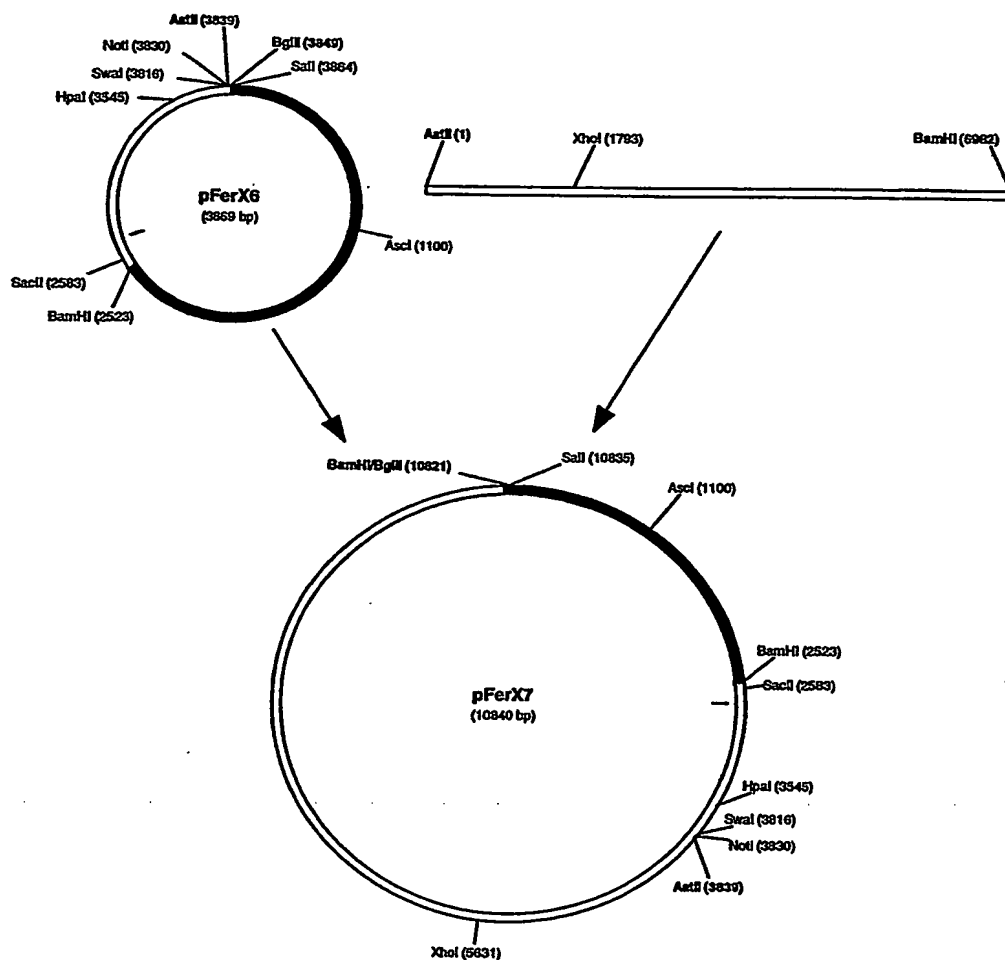


FIGURE 8

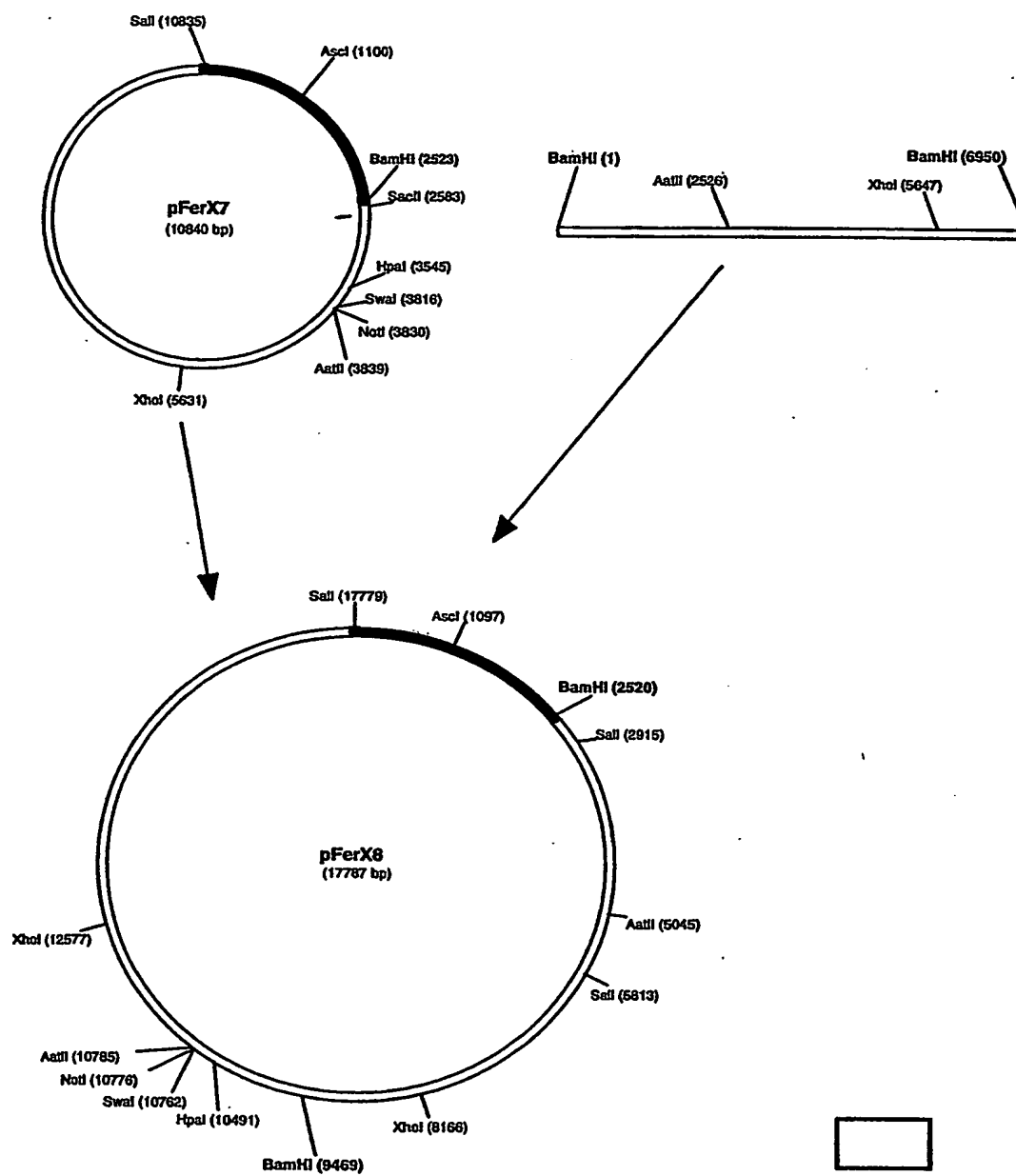


FIGURE 9

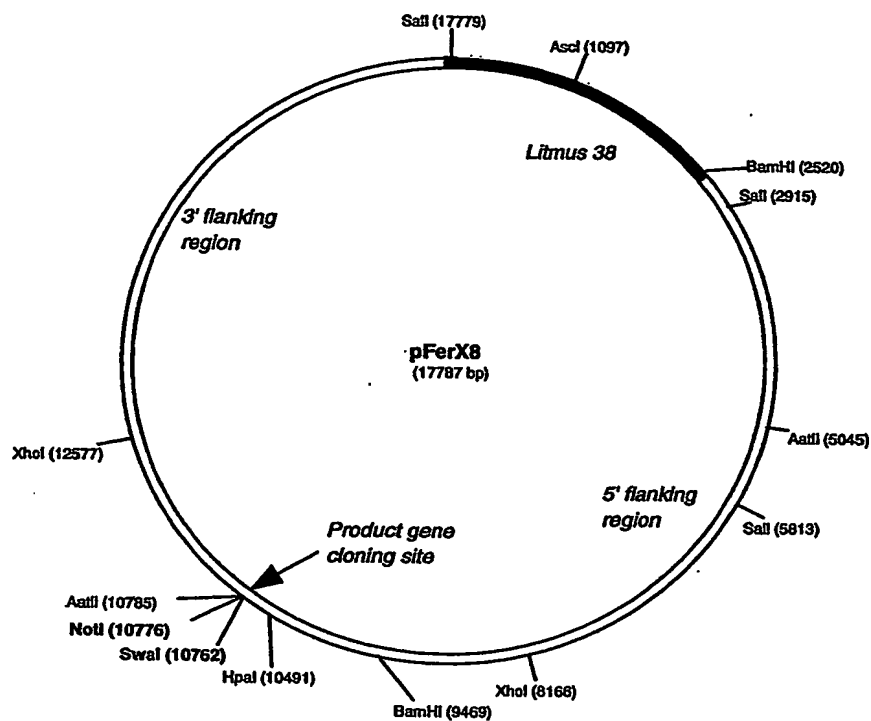


FIGURE 10

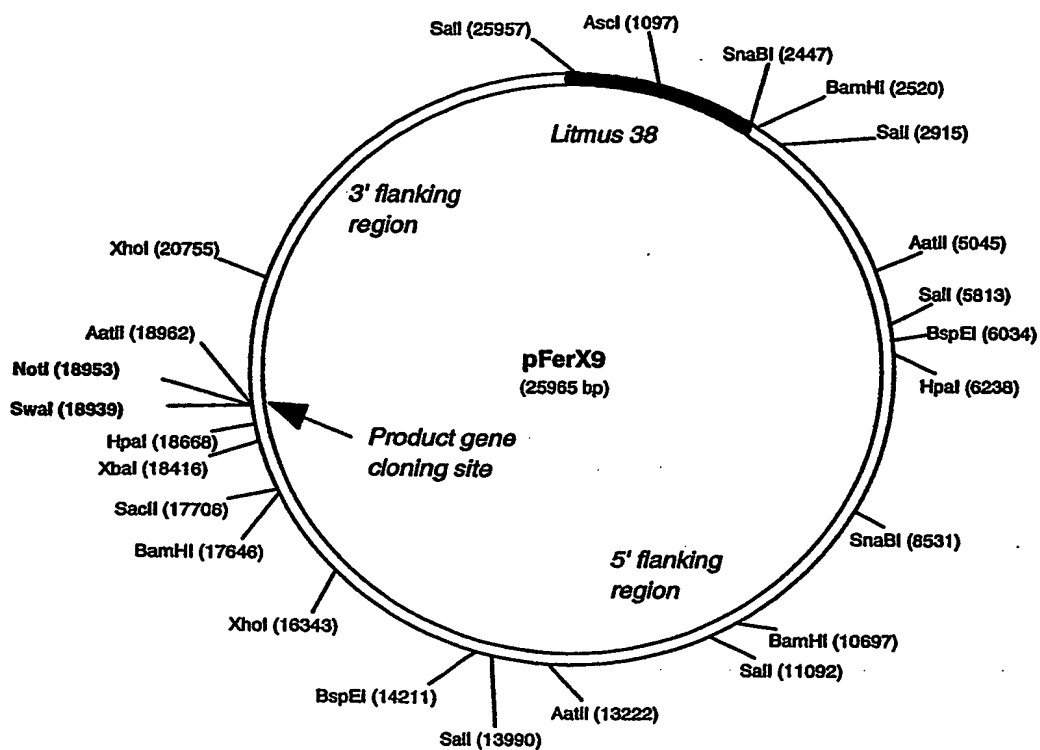


FIGURE 11

BamHI (9469) SacII
CAP (9521) KpnI (9533)
GGATCCCGCTATAAAGTGC GGCCCGCTGGTCCCTACGCCAGACGTTCTCGCCAGAGTCGCCGCGGTACCGGTGCTCG
ACCCCTCCGACCCCGTCCGGCCGCTTTGAGCCTGAGCCCTTTGCAACTTCGTGCTCCGCCGCTCCAGCGTCGCCTC
CGCGCCTCGCCAGCCGCCATC ATG gtgagtgcggcctggcctttggcggggcggaagagggtgcggcctggcct
►Met
cccttggggcacttgggtgagctggcggagggtgggttggggcgtggcctgctgcgggcttccccgccttccagcgcct
ttctggaaaatggagtttgtccggggttctttccaaaggcaggcagccctgccgtggcaagtctgagcacctagcgct
ttgtggctcctgcatagaccaggcagctcataacaccctgtgtttgaagccttagggctgtacaactgtcagcctctc
caatcaaccctgcagttaggtgcattttcctgcactctcgtccctccggtcacatggcctgcaggcttctctgtttg
gggtgtacatccagctccagttcctctgactatggcgggtctgcttgggtcatgggtgtggaatggcagccctggggcttg
gtacaaagaggcttatctctgtgaacttactctaaccacttctgaagcagcggcctctacatctctgcttatcacag
agcctcacttgcattgaaacttatcgctaggaatctcccttctgtaatcaccctgaccttgccaaggcatctagagt
actgtacgtttttaatttttatgtgcaccagttgttgcttactaacagaagtagtaggtaacatacttgttggaaaa
agccccaggttgggaaaaaacattatcgtggaatacaatacactgagtgccctaaaactgaaaatcaaagcttctcc
caatgtatttgtgctaaaaatacaatgccctcagttcttaaccaggtaatcagcagttggctgtctagctgaaaacctt
gagaccttgtgttaaccattttttttatttaacatgattgttgaaggagagaattgacctcccaatgtagggcacttt
agcacccccctctcagacaaatagatatggccttggcttaaagtttttctctgcactaatgtggagccatagaacc
cttgataaagccaagtcaccaagttgttttcccatccttactttaaggccaagtagggtgacaaacagcctttacca
AatII (10785)
ccattgcatctgccttgcgtgtgggatcaataacaaataccctttccatttAAATCTGCTAGCGGCCGCTGACGTCCC
CAAGGCCATGTGACTTTACTGGTCACTGAGGCAGTGTCATGTCATGTCAGGCTGCCTTTATCTTTCTATAAGTTGCACC
KpnI (10927)
AAAACATCTGCTTAAAAGTTCTTTAATTTGTACCATTTCTTCAAATAAAGAATTTTGGTACCCAGCTCTTGTGTGAT
TGAGGATGAGCGCACCAGCTTCCCTTGCCTCGGCTATACTAACCACACTGCA

FIGURE 12

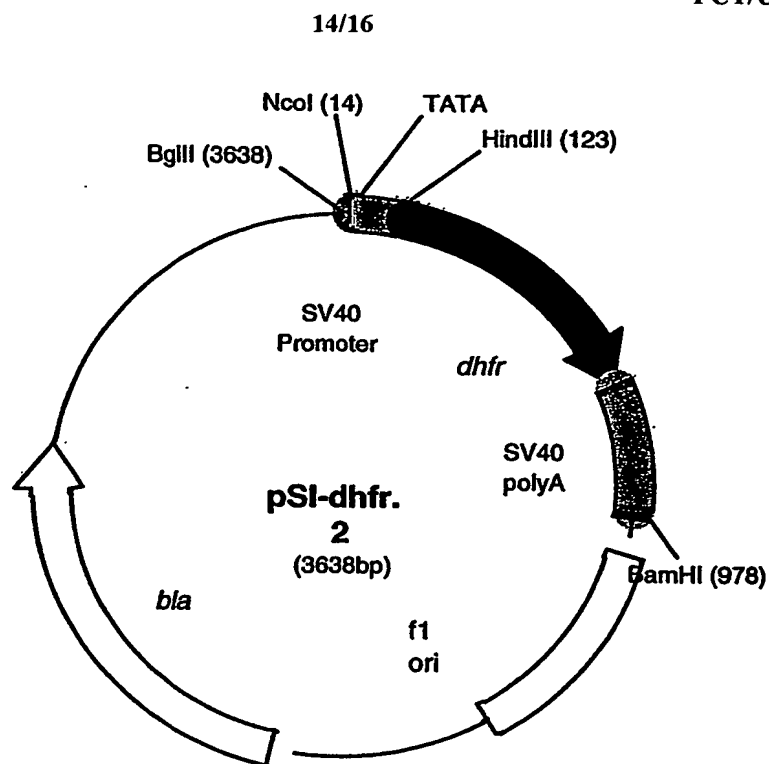
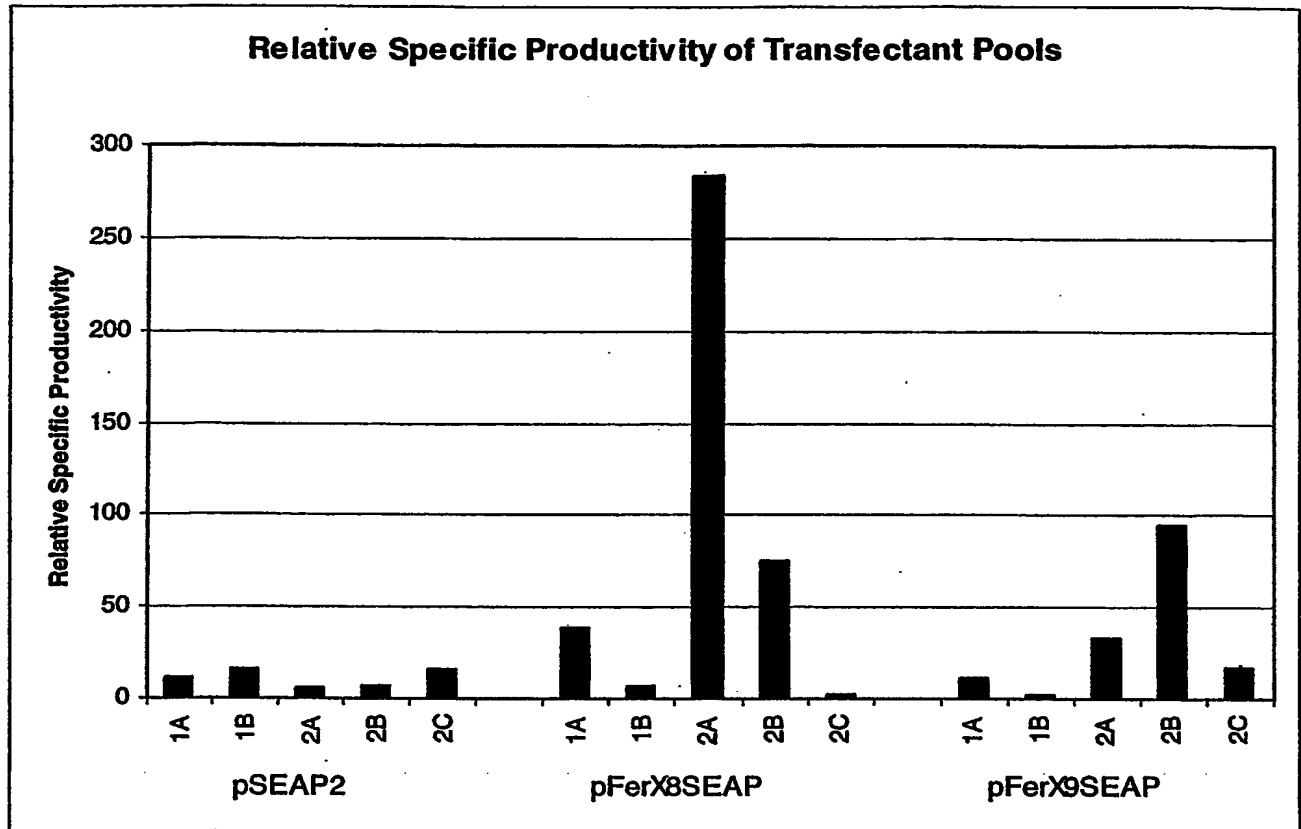


FIGURE 13

**FIGURE 14**

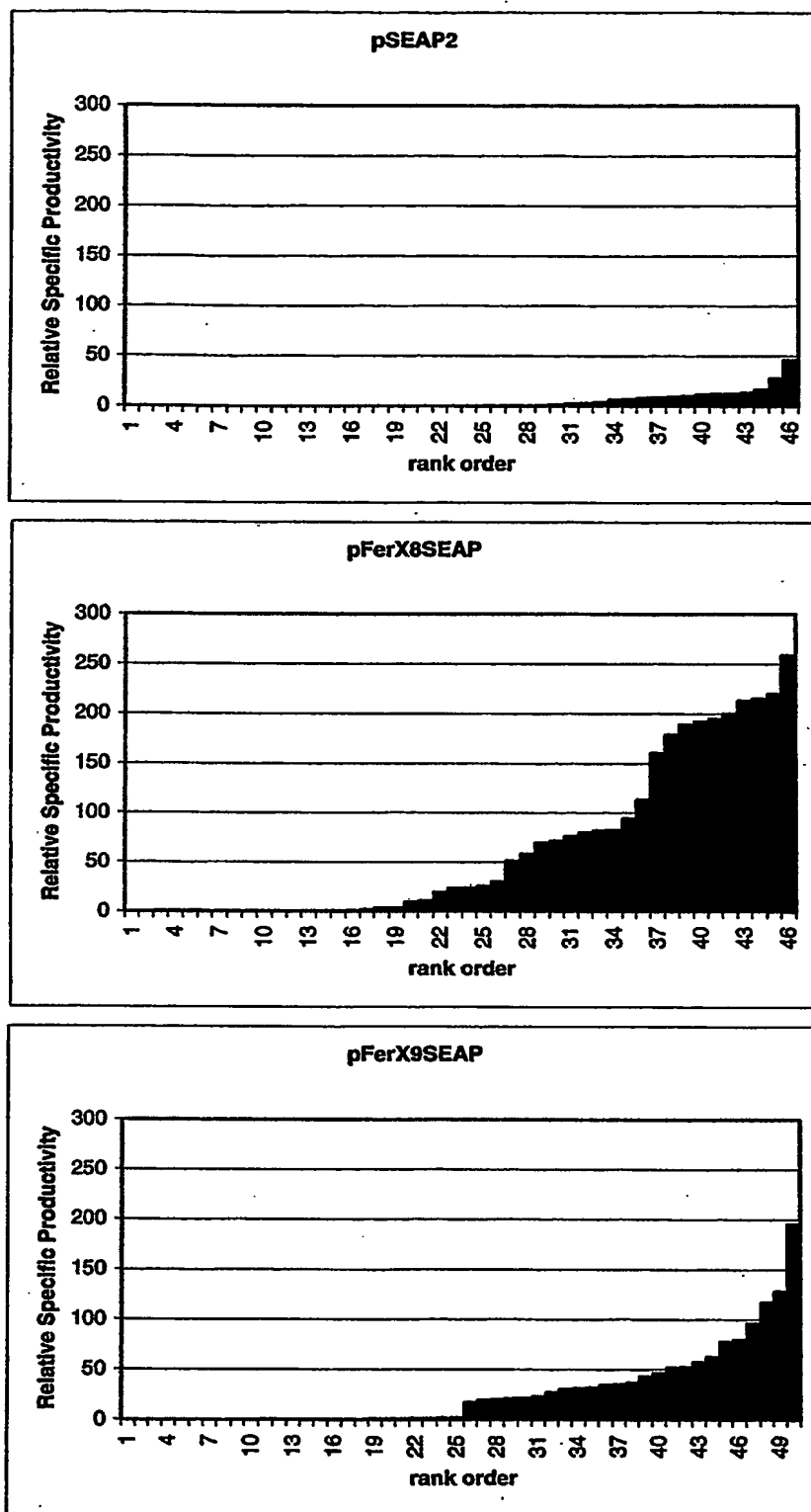


FIGURE 15

(19) World Intellectual Property
Organization
International Bureau



21 APR 2005

(43) International Publication Date
6 May 2004 (06.05.2004)

PCT

(10) International Publication Number
WO 2004/037982 A3

(51) International Patent Classification⁷: C12N 5/10,
15/85, 15/63, 15/64, C12P 21/00

SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(21) International Application Number:
PCT/US2003/033433

(22) International Filing Date: 22 October 2003 (22.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/421,252 24 October 2002 (24.10.2002) US

(71) Applicant (for all designated States except US): BIO-
GEN, INC. [US/US]; 14 Cambridge Center, Cambridge,
MA 02142 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): PRENTICE, Holly
[US/US]; 35 Nathan Lane, Carlisle, MA 01741 (US).

(74) Agent: FRASER, Janis, K., J., D.; Fish & Richardson
P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE,
GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR,
KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT,
RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted
a patent (Rule 4.17(ii)) for the following designations AE,
AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH,
CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES,
FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH,
PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN,
TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO
patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG,
ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU,
TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT,
RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

— as to the applicant's entitlement to claim the priority of the
earlier application (Rule 4.17(iii)) for the following design-
ations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY,
BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC,
EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU,
LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ,
OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY,
TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM,
ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ,
TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ,
MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY,
CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC,
NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG,
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

— with international search report

(88) Date of publication of the international search report:
30 September 2004

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: HIGH EXPRESSION LOCUS VECTOR BASED ON FERRITIN HEAVY CHAIN GENE LOCUS

(57) Abstract: High expression locus vectors based, in part, on the ferritin heavy chain locus are disclosed. The vectors include distal 5' flanking sequences and/or proximal 5' regulatory sequences derived from ferritin heavy chain locus. The vectors include a site for insertion of heterologous sequences and proximal 3' regulatory and distal 3' flanking sequences. The proximal 3' regulatory and distal 3' flanking sequences are optionally derived from the ferritin heavy chain locus. Cells transformed with the vectors, and methods of producing heterologous proteins encoded by the vectors, are also disclosed.

WO 2004/037982 A3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/33433

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 5/10, 15/85, 15/63, 15/64; C12P 21/00

US CL : 435/320.1, 69.1, 325, 455

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/320.1, 69.1, 325, 455

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| A | KWAK et al. Role for NF-KB in the Regulation of Ferritin by Tumor Necrosis Factor-alpha. The Journal of Biological Chemistry. June 1995, Vol. 270, No. 25, pages 15285-15293, especially pages 15288-15289. | 1-5, 18-19 |
| A | MARZIALI et al. Transcriptional Regulation of the Ferritin Heavy-Chain Gene: the Activity of the CCAAT Binding Factor NF-Y is Modulated in Heme-Treated Friend Leukemia Cells and during Monocyte-to-Macrophage Differentiation. Molecular and Cellular Biology. March 1997, Vol. 17, No. 3, pages 1387-1395, especially pages 1389-1390. | 1-5, 18-19 |
| A | SHIN et al. Enhanced Iron Uptake of Saccharomyces cerevisiae by Heterologous Expression of a Tadpole Ferritin Gene. Applied and Environmental Microbiology. March 2001, Vol. 67, No. 3, pages 1280-1283, especially page 1282. | 1-5, 18-19 |
| A | GUO et al. Expression and Loading of Recombinant Heavy and Light Chain Homopolymers of Rat Liver Ferritin. Archives of Biochemistry and Biophysics. November 1996, Vol. 335, No. 1, pages 197-204, especially pages 202-203. | 1-5, 18-19 |



Further documents are listed in the continuation of Box C.



See patent family annex.

| | |
|---|--|
| * Special categories of cited documents: | |
| "A" document defining the general state of the art which is not considered to be of particular relevance | "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "E" earlier application or patent published on or after the international filing date | "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "O" document referring to an oral disclosure, use, exhibition or other means | "&" document member of the same patent family |
| "P" document published prior to the international filing date but later than the priority date claimed | |

Date of the actual completion of the international search

30 May 2004 (30.05.2004)

Date of mailing of the international search report

02 JUL 2004

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Facsimile No. (703) 305-3230

Authorized officer

David Guzo

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/33433

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claim Nos.: 6-17, 20
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US03/33433

Continuation of B. FIELDS SEARCHED Item 3:

WEST, Dialog, Medline, Biotech, Biosci

Search terms: ferritin heavy chain, transcriptional initiation, expression vector, promoter, 5' regulatory sequence

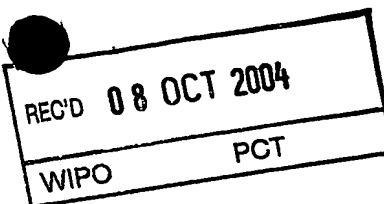
PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

21 APR 2005



| | | |
|---|---|---|
| Applicant's or agent's file reference 13751-021WO1 | FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416) | |
| International application No. PCT/US03/33433 | International filing date (day/month/year) 22 October 2003 (22.10.2003) | Priority date (day/month/year) 24 October 2002 (24.10.2002) |
| International Patent Classification (IPC) or national classification and IPC IPC(7): C12N 5/10, 15/85, 15/63, 15/64; C12P 21/00 and US Cl.: 435/320.1, 69.1, 325, 455 | | |
| Applicant BIOGEN, INC. | | |

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.
2. This REPORT consists of a total of 4 sheets, including this cover sheet.
☐ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of ___ sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☒ Non-establishment of report with regard to novelty, inventive step and industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☐ Certain observations on the international application

| | |
|---|--|
| Date of submission of the demand 24 May 2004 (24.05.2004) | Date of completion of this report 19 September 2004 (19.09.2004) |
| Name and mailing address of the IPEA/US Mail Stop PCT, Attn: IPEA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230 | Authorized officer David Guzo Telephone No. (571) 272-1600 |

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/US03/33433

I. Basis of the report**1. With regard to the elements of the international application:***

- ☒ the international application as originally filed.
- ☒ the description:
pages 1-29 as originally filed
pages NONE, filed with the demand
pages NONE, filed with the letter of _____.
- ☒ the claims:
pages 30-33, as originally filed
pages NONE, as amended (together with any statement) under Article 19
pages NONE, filed with the demand
pages NONE, filed with the letter of _____.
- ☒ the drawings:
pages 1-16, as originally filed
pages NONE, filed with the demand
pages NONE, filed with the letter of _____.
- ☐ the sequence listing part of the description:
pages NONE, as originally filed
pages NONE, filed with the demand
pages NONE, filed with the letter of _____.

2. With regard to the language, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

These elements were available or furnished to this Authority in the following language _____ which is:

- ☐ the language of a translation furnished for the purposes of international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of the translation furnished for the purposes of international preliminary examination (under Rules 55.2 and/or 55.3).

3. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☐ contained in the international application in printed form.
- ☐ filed together with the international application in computer readable form.
- ☒ furnished subsequently to this Authority in written form.
- ☒ furnished subsequently to this Authority in computer readable form.
- ☒ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☒ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. ☐ The amendments have resulted in the cancellation of:

- ☐ the description, pages NONE
- ☐ the claims, Nos. NONE
- ☐ the drawings, sheets/fig NONE

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).**

* Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17).

** Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/US03/33433

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The question whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), or to be industrially applicable have not been and will not be examined in respect of:

- ☐ the entire international application,
☒ claims Nos. 6-17 and 20

because:

- ☐ the said international application, or the said claim Nos. _____ relate to the following subject matter which does not require international preliminary examination (*specify*):

- ☒ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. 6-17 and 20 are so unclear that no meaningful opinion could be formed (*specify*):

Claims 6-17 and 20 are not examined because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

- ☐ the claims, or said claims Nos. _____ are so inadequately supported by the description that no meaningful opinion could be formed.
☐ no international search report has been established for said claims Nos. _____

2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

- ☐ the written form has not been furnished or does not comply with the standard.
☐ the computer readable form has not been furnished or does not comply with the standard.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.
PCT/US03/33433

V. Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. STATEMENT

| | | |
|-------------------------------|------------------------------|-----|
| Novelty (N) | Claims <u>1-5, 18 and 19</u> | YES |
| | Claims <u>NONE</u> | NO |
| Inventive Step (IS) | Claims <u>1-5, 18 and 19</u> | YES |
| | Claims <u>NONE</u> | NO |
| Industrial Applicability (IA) | Claims <u>1-5, 18 and 19</u> | YES |
| | Claims <u>NONE</u> | NO |

2. CITATIONS AND EXPLANATIONS

Claims 1-5 and 18-19 meet the criteria set out in PCT Article 33(2) and (3), because the prior art does not teach or fairly suggest genetic vectors with distal 5' sequences flanking a heterologous coding sequence, wherein said 5' sequences comprise a sequence of at least 100 bases having at least 70% identity to a nucleotide sequence found between 20 bp and 100,000 bp of a transcriptional start site of a ferritin heavy chain or with proximal 5' regulatory sequences comprising a sequence of at least 20 bp having at least 70% identity to a nucleotide sequence found between 1 bp and 10,000 bp 5' of a translational initiation codon of a ferritin heavy chain locus.

Claims 1-5 and 18-19 meet the criteria set out in PCT Article 33(4), and thus possess industrial applicability because the subject matter claimed can be made or used in industry. Specifically, the claimed vectors can be used to stably express high levels of pharmaceutically important proteins in cells in culture.

PATENT COOPERATION TREATY

PCT

NOTE OF INFORMAL COMMUNICATION WITH THE APPLICANT

(PCT Rule 66.6)

| | | |
|--|--|---|
| International application No. PCT/US03/33433 | Applicant's or agent's file reference 13751-021WO1 | Date of informal communication (day/month/year) 13 September 2004 (13.09.2004) |
| Applicant BIOGEN, INC. | | |

| | | |
|---|---|--|
| <u>Communication</u> <input checked="" type="checkbox"/> by telephone <input type="checkbox"/> personal | <u>Participants</u> <input checked="" type="checkbox"/> Applicant: Biogen, Inc. <input checked="" type="checkbox"/> Agent: Jack Brennan <input checked="" type="checkbox"/> Examiner(s): David Guzo | <input checked="" type="checkbox"/> Identity checked <input checked="" type="checkbox"/> authorization checked <input type="checkbox"/> personally known |
|---|---|--|

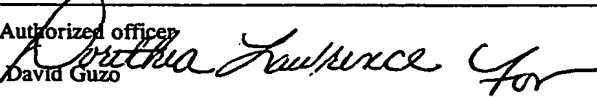
Summary of communication:

Applicants' representative gave approval for skipping the written opinion (408) and going to the preliminary examination (409) because no prior art has been applied against the claims.

☐ An extension of time limit is granted (Form PCT/IPEA/427).

☒ A copy of this note is being sent to the applicant with Form ~~PCT/IPEA/429~~.

PCT/IPEA/424.

| | |
|---|--|
| Name and mailing address of the IPEA/US Mail Stop PCT, Attn: IPEA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230 | Authorized officer  David Guzo Telephone No. (571) 272.1600 |
|---|--|

Form PCT/IPEA/428 (July 1992)